nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

QUANTUM LOGIC

A two-qubit exchange gate based on entangled electron spins on atoms

PAGE 371

PUBLISHING

A MINE OF INFORMATION

The data store enabling text analysis of research papers

PAGE 315

BIOTECHNOLOGY

SKIN PATCHES & SMART TATTOOS

Next steps for health-checking biosensors

PAGE 319

EVOLUTION

DEVELOPMENT OF A SEA SQUIRT

Gene expression in embryos maps stages of growth

PAGES 333 & 349

NATURE.COM

18 July 2019

Vol. 571, No. 7765

THIS WEEK

EDITORIALS

WORLD VIEW Bolster the data to build the foundations for sustainable development **p.299**

COOL TACTICS Minty menthol acts as switch in genetic circuits **p.300**

FACE OF THE FIFTY Alan Turing will grace Bank of England's new £50 note **p.303**

Balance security with openness

As US federal agencies tighten their anti-espionage policies, universities must protect their Chinese and Chinese American communities from profiling.

As the president of one of the world's most diverse and international universities, Rafael Reif probably never expected to be writing to students and staff explaining why immigration and international collaboration are not a threat to the security and integrity of the United States.

Reif, head of the Massachusetts Institute of Technology (MIT) in Cambridge, wrote an open letter in praise of diversity last month, after Chinese and Chinese American members of the institute's community reported being interviewed by law-enforcement agencies and asked about their links to China.

Targeting individuals from particular ethnicities in this way is not acceptable. As Venki Ramakrishnan, president of the UK Royal Society, writes on page 326, it violates the principle of "innocent until proven guilty", an axiom of modern democracies. MIT's Office of the General Counsel should not have had to organize a special briefing, as it did last week, giving leading scientists from the affected communities contact details for MIT lawyers and advising them on what to do if approached, for example, by the FBI.

This troubling situation began before the last US presidential election, when lawmakers became concerned about academia's global ties, and in particular that its links to China represent an open door to spying and intellectual-property theft. But at the same time, Barack Obama's administration saw benefits from collaborations between the best US and Chinese researchers — funded by their respective governments, no less. But China's continued aspirations to become a global power, and the election of President Donald Trump have now escalated concerns.

Threats to the United States from China's government and its companies were among the top discussion items at last September's FBI Academic Summit — the annual gathering at which university and research leaders discuss security with representatives from law-enforcement, intelligence and security agencies. Meanwhile, in Congress, much of the charge continues to be led by Charles Grassley, chair of the Senate finance committee and Republican senator from Iowa.

GROWING SCRUTINY

Grassley's committee has been in correspondence with the US National Science Foundation (NSF) and the National Institutes of Health (NIH), demanding to know how they vet their grant recipients, how much they spend on policing misconduct, and the measures being taken to "punish foreign agents". Even though funding agencies are not in the business of punishment, this scrutiny has unsettled both agencies. But it has also exposed gaps in how they monitor compliance with grant policies. This is prompting them to take action.

After a year-long sweep, the NIH says it has uncovered 180 scientists at 60 grant-holding institutions who have either violated peer-review policies — for example by sharing grant proposals with representatives of non-US institutions — or have failed to disclose non-US financial links, often with institutions in China. Not all of this will be espionage-related, but if the NIH is correct, it validates director Francis Collins's decision to write to 10,000 universities last year, reminding them of their disclosure obligations.

More problematic are some of the recommendations from a group of university leaders commissioned by the NIH to advise the agency on tackling what it calls foreign influences on research integrity. The panel's advice includes updating NIH conflict-of-interest policies so that funded investigators must disclose any work with an international partner that might overlap with the scope of an agency award.

But the panel also recommends that universities vet academic staff before hiring them — implying that this would be an extra security check. It suggests that universities increase scrutiny of the movements of overseas visitors, and that they conduct written interviews with scientists who have travelled to "select countries" on research trips. University leaders, meanwhile, are encouraged to boost their awareness of "scientific topics that are more prone to interest by untoward actors".

"The response to scrutiny needs to be both proportionate and evidence-based."

Universities and the NIH are also being encouraged to carry out a "broad education campaign" to reinforce the importance of adhering to NIH policies. This also includes asking investigators to keep records of interviews with students and postdocs, concerning their plans after leaving a laboratory.

Although no ethnic group is named, the panel mentions that China is a country of concern. This is fuelling fear among Chinese and Chinese American academics that their ethnicity and their scientific work are the main target — especially for those in cutting-edge fields. The NIH denies this (see *Nature* **571**, 157; 2019).

So far, these are recommendations, not requirements. However, universities may well be minded to comply, considering that many of the same questions are being asked by the Senate finance committee, and that not doing so could risk their future NIH funding.

At the same time, both the NIH and the NSF know all too well that the shape and scale of their response to the scrutiny need to be both proportionate and evidence-based. And what little evidence there is from other countries indicates that broad academic monitoring exercises, for example those in counter-terrorism that target students from minority communities, have questionable value for reducing security risks.

The NSF, in its own response published last week, says that it has commissioned the independent scientific advisory group JASON to assess how universities could maintain the balance between openness and security. This is the kind of response that others should adopt, although 'security' must include the security of those people from minority groups, such as Chinese and Chinese American scientists, who have made an important contribution to the success of US science.

Every academic institution has a duty to prioritize the safety and security of its communities, but it must also, as Reif reminded his colleagues last month, "take great care not to create a toxic atmosphere of unfounded suspicion and fear". ■

VICKY SHARP PHOTOGRAPHY

The missing ingredient for a better world: data

Unless governments establish competent monitoring systems, the world will not reach the UN Sustainable Development Goals, says **Jessica Espey**.

In 2013, I worked in Liberia's Ministry of Finance and Development Planning. My office was in a run-down beachside building with intermittent electricity and water. One day, the generator surged. Within seconds, we smelt singed plastic. Our computers, and other equipment the government could ill afford to replace, were ruined. The damage at the national statistics office next door was devastating. Reams of survey data typed in from paper reports were lost, along with tens of other data sets about educational outcomes, poverty rates and access to services. They had all been saved on just one computer.

Cash-strapped, infrastructure-limited national data systems run by staff who lack training and authority are common among poor countries. They are the biggest barrier to achieving the Sustainable Development Goals (SDGs) — covering everything from cleaner water to fairer societies — set by 193 countries and the United Nations in 2015, meant to put the world on a path to a sustainable future by 2030. As a forum to consider progress on SDGs meets this week, it must consider this fact: none of these goals can be met without a data revolution.

Many national statistics are compiled on paper, manually inputted to old computers, and unavailable or inconsistently accessible online. Thus, government statistics are not referred to for day-to-day (or even week-to-week) decisions. Those data that are available are usually out of date: only 35% of sub-Saharan countries have poverty data that were updated since 2015.

The creation of fit-for-purpose systems will require a massive, coordinated commitment from governments and the international community. The research group I lead in the UN Sustainable Development Solutions Network outlines the needs in a forthcoming report (see www.sdsntrends.org). Four building blocks are crucial: strong governance; appropriate policies and standards; a culture of innovation; and a case that can convince global donors and national governments to make sufficient investments.

Skilled staff working on innovative, transparent data systems could provide high-quality, timely data to show environmental changes, social conditions and economic fluctuations. In pilot programmes in West Africa in 2014, telecommunications companies and governments teamed up to track population movement and disease spread. These data could then be used to set up countermeasures, such as community education and washing stations, in areas where they would be most effective.

With daily Earth observation data, governments could monitor erosion, sand mining and illegal development and then act to stabilize fragile coastlines. Interconnected administrative systems could help to give vulnerable people access to health facilities, social services and entitlements. Data systems are the mortar with which a sustainable planet and society will be built.

DATA SYSTEMS ARE THE MORTAR WITH WHICH A SUSTAINABLE PLANET AND SOCIETY WILL BE BUILT.

Governments first need to appoint and empower the right people. Particularly important is a chief statistician mandated to work across government agencies to promote data availability, and advocate for the use of data. The Philippines and New Zealand are leading the way by consolidating relevant functions under a central authority and appointing coordinators to negotiate data-sharing agreements across and beyond government departments.

The UN Statistical Commission, founded in 1947, needs reform to support such efforts. It could start by being more inclusive, recruiting new data providers and building trust and common cause among them. For example, civil society groups, scientists and private data providers should be invited to the formal commission proceedings.

Data collectors need clear standards, policies and terminology. These can make or break governments' and private companies' will to collaborate and support a shared mission. For example, the lack of consensus on definitions for natural and human-made hazards means that governments, the UN, researchers, insurance companies and other private entities struggle to share information, despite having committed to doing so under the SDGs, the Paris climate agreement and other international pacts, such as the Sendai Framework for Disaster Risk Reduction. Thus, governments cannot accurately measure how many people have been affected by a hurricane or a tsunami, and the UN does not know how much assistance to send.

We need to move towards a 'digital ecosystem' that encourages contributions — from citizen science, national institutions and transnational corporations — and collaboration. The goal should be counting everyone and proving everyone counts.

Investing in data generates huge economic, social and environmental returns. Earth observation data from the NASA and US Geological Survey Landsat satellites produce an estimated benefit of US$2.19 billion a year from applications such as smarter land-use planning and more timely responses to natural disasters.

Governments and international donors must boost their investments and make better use of existing resources, in part by establishing ways to coordinate resources that leave no country or region behind.

Four years have elapsed since leaders committed to achieving the SDGs in their countries by 2030. Eleven years remain. Sustainable development will falter without data. We must put national data systems in place, or the SDGs will be little more than feel-good aspirations. ■

Jessica Espey *directs TReNDS, a data-focused knowledge network under the UN Sustainable Development Solutions Network.*
e-mail: jessica.espey@unsdsn.org

SEVEN DAYS *The news in brief*

FACILITIES

Stem-cell agency

The California Institute for Regenerative Medicine (CIRM) in Oakland has run out of money for new projects and will no longer accept grant applications. CIRM was created in 2004 after California voters approved a ballot initiative that provided US$3 billion for stem-cell research. Over the past 15 years, CIRM has funded more than 1,000 research projects, including 55 clinical trials. But that money has dried up. The agency has just $33 million left for new grants, yet it is evaluating a backlog of applications that would cost $88 million. The 260 or so projects that CIRM already funds will continue as planned, but its dwindling coffers cast doubt over the future of stem-cell research in California. The advocacy group Americans for Cures, founded by CIRM supporters, is considering spearheading a 2020 ballot initiative to provide $5.5 billion for the struggling agency, but has not announced concrete plans.

Telescope project

An attempt to restart construction of the controversial Thirty Meter Telescope (TMT) in Hawaii is being delayed by protests. Previous protests and legal challenges had postponed the TMT team's plans to build atop Mauna Kea, a mountain sacred to Native Hawaiians, over the past four years. On 10 July, Hawaii governor David Ige said that construction would restart on 15 July — prompting another round of protests. TMT officials had estimated the project's cost to be US$1.4 billion, but that ffigure has almost certainly risen since April 2015, when protests halted the first attempt to build the telescope. The Hawaii supreme court revoked the TMT's construction permit later that year, sending the project to Hawaii's Board of Land and Natural Resources to undergo a second permit process. Last October, the court ruled that the telescope's second permit was valid.

ROSCOSMOS

X-ray space telescope blasts off

A German–Russian space telescope that will produce the first full-sky maps in high-energy X-rays has left Earth. Spectrum-Roentgen-Gamma (SRG) lifted off on a Russian-built Proton-M rocket on 13 July from the Baikonur Cosmodrome in Kazakhstan. It is now on a 100-day journey to L2, the second Lagrangian point, a gravitationally stable parking spot that trails Earth's orbit around the Sun. From there, SRG's two on-board telescopes — one built by a Russian team and the other by a German one — will chart maps of the Universe's 'hard' X-rays, which have energies of up to about 30 kiloelectronvolts. It is expected to detect up to 100,000 galactic clusters, 3 million supermassive black holes and X-rays from as many as 700,000 stars in the Milky Way. SRG's main scientific goal is to investigate the nature of dark energy, the mysterious force that is accelerating the Universe's expansion.

EVENTS

Ebola outbreak

The Ebola outbreak in the Democratic Republic of the Congo (DRC) has spread to Goma, one of the country's biggest cities. On 14 July, the DRC health ministry said that an evangelical pastor had been diagnosed with the disease at a treatment centre there. The pastor had preached in Butembo, a hotspot of the outbreak, where he laid hands on people who were ill. After he began to feel sick, he travelled to Goma by bus — a two-day journey. Doctors there transferred him to an Ebola treatment centre. The ministry has given an Ebola vaccine to all passengers on the bus and will monitor them for 21 days. It says that the risk of Ebola spreading in Goma is low because the pastor was diagnosed quickly. Ebola has killed an estimated 1,665 people since 1 August 2018, according to the DRC health ministry.

Satnav troubles

Galileo, the European Union's global satellite-navigation system, has experienced a continuous service outage since 11 July. The navigation and timekeeping services that it offers users worldwide had not been restored as *Nature* went to press. Galileo officials said that a technical issue in its ground infrastructure — the cause of which has been identified — is behind the outage, but released few further details. The

ground-tracking network is a crucial part of the satnav system, and uploads orbit corrections to the 26 Galileo satellites in near real-time. Galileo began offering services in 2016. A 15 July statement on the Galileo website emphasized that the system is still in the pilot phase, which is designed to detect technical issues before it becomes fully operational next year.

PEOPLE

Turing honoured

Mathematician Alan Turing will adorn the Bank of England's new £50 note, beating physicist Stephen Hawking and a legion of other renowned scientists considered for the honour, the bank announced on 15 July. Turing is considered the founder of computer science and a national hero, having helped to crack the code of German encryption machines during the Second World War. Turing was gay, and was convicted for gross indecency in the 1950s. He died in 1954, aged 41, and was pardoned by the Queen in 2013. In 2018, the bank called on the public to suggest scientists who could appear on a redesigned £50 note, and received nominations for 989 eligible individuals. A committee shortlisted 12, and the bank's governor, Mark Carney (**pictured**), made the final decision. Palaeontologist Mary Anning, crystallographer Rosalind Franklin and mathematician Srinivasa Ramanujan were among the contenders.

OLI SCARFF/GETTY

SPACE

Mission aborted

India's space agency scrapped the launch of its second Moon mission 56 minutes before scheduled lift-off on 15 July because of a technical hitch with its rocket's engine. The Chandrayaan-2 spacecraft will head for the Moon's south pole, and India hopes that it will become its first mission to land on the lunar surface. The spacecraft is equipped with 14 instruments — 13 Indian and one from NASA — and aims to study the Moon's rocks, soil, minerals and atmosphere, and to detect water. The craft was set to launch on India's Geosynchronous Satellite Launch Vehicle Mark III. Engineers had filled the rocket's cryogenic engine with fuel, but postponed the launch as a precautionary measure after detecting a leak. The Indian Space Research Organisation (ISRO) said that it might announce a new launch time after analysing the rocket. Chandrayaan-2 comes 11 years after India's maiden lunar mission in 2008, which orbited the Moon. ISRO announced plans for Chandrayaan-2 in 2008. Russia pulled out of a planned collaboration on the spacecraft in 2013.

RESEARCH

Mosquito success

A trial to control the Asian tiger mosquito (*Aedes albopictus*), the world's most invasive mosquito species, has nearly eradicated the insects from two sites in Guangzhou, China. For the first time, researchers field-tested a combination of two techniques that have been used separately in previous mosquito elimination trials. The two-pronged approach involves infecting males with a strain of *Wolbachia* bacteria, hindering the insects' ability to transmit disease-causing viruses such as dengue and Zika, and preventing them from producing offspring with females without the same strain. Researchers then gave the female mosquitoes low-dose radiation to sterilize them. After the researchers released millions of the mosquitoes in 2016 and 2017, the overall number of wild adult mosquitoes at the two test sites decreased by up to 94% each year. The researchers published their results on 17 July (X. Zheng *et al. Nature* https://doi.org/10.1038/s41586-019-1407-9; 2019).

TREND WATCH

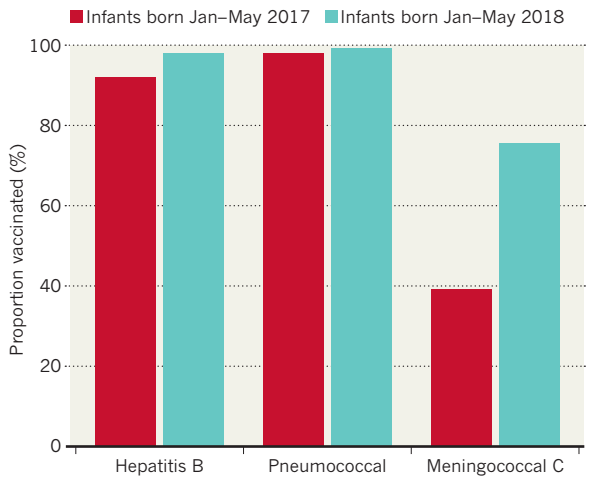
Vaccination coverage in France and Italy has increased following the expansion of mandatory vaccination laws, according to two analyses. Both countries already required children to receive diphtheria, tetanus and polio vaccinations. Italy also mandated immunization against hepatitis B. France — home to one of the highest rates of vaccine mistrust in Europe — made eight more vaccines mandatory for babies born from 2018 onwards, including vaccines against hepatitis B, pneumococcal and meningococcal C diseases.

A preliminary analysis of the French law's impact found that rates of these vaccinations were higher among babies born in the first five months of 2018 than among those born in the same period in the previous year.

Italy made another six vaccines mandatory in July 2017 after a large outbreak of measles that affected more than 4,000 Italians. The second paper found that coverage of these immunizations increased: in 2016, 87% of 2-year-olds were vaccinated against measles, but by mid-2018, that number had risen to 94% among 30-month-old children. The studies were published in June in *Eurosurveillance*.

MANDATORY VACCINES

After France expanded its mandatory vaccination programme in 2018, the proportion of children under one year old who received the newly compulsory immunizations increased.



SOURCE: D. LÉVY-BRUHL *ET AL. EURO SURVEILLANCE* **24**, 1900301 (2019).

NEWS IN FOCUS

MAX PLANCK INSTITUTE OF MOLECULAR CELL BIOLOGY AND GENETICS

Suzanne Eaton was a developmental biologist at one of Germany's prestigious Max Planck institutes.

RESEARCHER DEATH

Scientists pay tribute to biologist killed in Crete

Suzanne Eaton's body was found last week near the site of a meeting she had been attending.

BY ALISON ABBOTT

Scientists, friends and family members have expressed their shock and grief over the death of developmental biologist Suzanne Eaton, who was killed earlier this month on the Greek island of Crete, where she had been attending a scientific conference.

Greek police said in a statement posted online on 16 July that a 27-year-old Greek man who was questioned as the main suspect in the homicide "admitted his guilt and today he will be brought to justice".

Eaton's body was discovered in cave on 8 July, several days after she went missing after going out for a run. Coroner Antoni Papadomanolakis, who is investigating Eaton's death, told the German television broadcaster RTL on 13 July that the cause of her death was suffocation.

One conference attendee has also described to *Nature* what happened when researchers realized that she was missing.

Eaton, who was 59 and a researcher at one of Germany's prestigious Max Planck institutes, was last seen on 2 July. She was in Greece for a conference on insect hormones at the Orthodox Academy of Crete in Kolymbari, a coastal village in the northwest of the island. Her body was found after a ▶

five-day search, 10 kilometres from the conference venue, and police opened a homicide investigation.

On a web page posted last week by the Max Planck Institute of Molecular Cell Biology and Genetics in Dresden, Germany, where Eaton worked, colleagues describe her as a world-renowned developmental biologist of singular passion, depth and breadth.

Eaton studied how particular molecules control embryonic development in fruit flies, and she had been scheduled to give the meeting's plenary lecture two days after the date of her disappearance.

'DEEP, DEEP SORROW'

"Her curiosity and enthusiasm for discovery was infectious," write her lab members. "She was our leader, our role-model, our mentor, our friend," they say. "Her sudden and tragic death has left us stunned and enveloped in deep, deep sorrow."

Colleagues also praised the work–life balance she achieved — she was the mother of two boys, as well as a talented musician and a black belt in tae kwon do.

"She worried that it was impossible to give both her science and her family her all," writes her sister. "With a deep sensitivity and compassion, she somehow made us all a priority."

Meeting attendees were thrown into turmoil when they realized Eaton was missing, says François Leulier, a molecular geneticist at the Institute of Functional Genomics of Lyons, France, who was also a conference plenary speaker. They had seen Eaton playing the piano on the afternoon that she disappeared, and thought little of the fact that she did not attend the evening session that day, he says.

"With her extensive scientific culture, she brought a richness to every meeting."

There were no lectures the following afternoon and some attendees, including the meeting organizers, began to discuss Eaton's absence. "We hoped that she had joined the conference excursion, but when they returned for the evening session without her we were really worried," says Leulier.

The group of attendees went to her room and found the wake-up alarm on her smartphone still ringing, indicating that she had not been there overnight. They drove straight to the police to report her missing, Leulier says, then at daybreak they divided into search groups and began to comb the shore and trails. Later that day, the police asked the scientists to remain at the conference centre to allow them to take charge of the search.

Leulier says he knew Eaton from the international conference circuit, where she was renowned for driving lively discussions. "With her extensive scientific culture, she brought a richness to every meeting," he says. "She asked probing questions on every subject in such a subtle, empathic and positive way."

During her career, Eaton had also been a staff scientist at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany. "The EMBL community is in shock and mourning," says a tribute web page posted by the laboratory. Cell biologist Kai Simons, Eaton's mentor at EMBL, writes that "she represented a modern Renaissance scientist in the sheer scope of her activities". ■

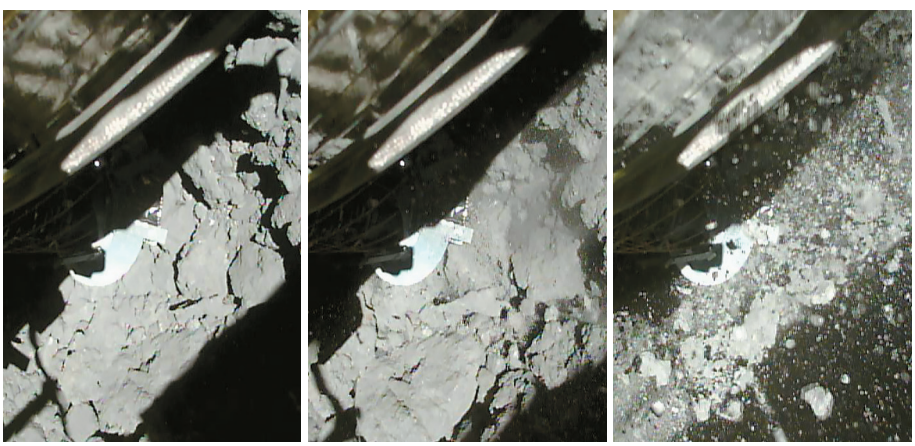
SOLAR SYSTEM

Japanese spacecraft probes asteroid's guts for first time

Hayabusa2 touched down on Ryugu to collect material from beneath the surface.

BY DAVIDE CASTELVECCHI

Japan's Hayabusa2 asteroid mission has performed the last major act in its saga of space exploration. At 10:18 a.m. Tokyo time on 11 July, the spacecraft descended on the asteroid Ryugu for the second time this year, to collect material from a crater it gouged out in April by striking the body's surface with a pellet. If the collection was successful — something that the mission team will not know for a while — it will be the first time in history that a mission has gathered material from an asteroid's innards.



JAXA

Images taken by Hayabusa2 as it descended towards the asteroid Ryugu.

The probe collected a sample from Ryugu's surface in February. After it returns its booty to Earth next year, scientists will be able to compare the composition of material from the two touchdown sites. That could reveal how exposure to the rigours of space, and in particular solar heating, solar wind and cosmic rays, affected the chemistry on the surface.

"This is a cornucopia of a mission," says Lucy McFadden, a planetary astronomer at NASA's Goddard Space Flight Center in Greenbelt, Maryland.

Hayabusa2 arrived at Ryugu in June 2018. It deployed landers that took magnetic, chemical and other measurements and sent pictures back. The spacecraft completed its first touchdown in February this year and then, in April, it shot a projectile that produced a 10-metre-wide crater, uncovering material under the asteroid's surface. Later this year, Hayabusa2 will turn back to Earth, where by the end of 2020 it is expected to deliver its samples for analysis.

In its latest move, Hayabusa2 aimed for a spot just outside the crater, rather than

descending inside the crater itself, which would have been "rather risky", mission manager Makoto Yoshikawa of the Japan Aerospace Exploration Agency (JAXA) Institute of Space and Aeronautical Science in Sagamihara told *Nature*.

"If you're going into a depression, then you have to worry about things like the solar panels sticking out" and potentially colliding with the surface, says Harold Connolly, a cosmochemist at Rowan University in Glassboro, New Jersey, and a co-investigator on the mission team. He is also working on NASA's OSIRIS-REx mission, which is exploring a similar body — called Bennu — and plans to collect material from its surface next year. The two missions exchange information and collaborate, in part by sharing staff.

The 1-kilometre-wide Ryugu is what scientists call a rubble-pile asteroid: a collection of rocks and dust held together loosely by gravity. Its low density — only slightly higher than that of liquid water — suggests that it is mostly empty space, and that it has accumulated from debris produced by a collision of other bodies, Connolly says.

Suction does not work in the vacuum of space, and Ryugu has almost no gravity. So the team devised an original technique that allows the spacecraft to pick up material while bouncing on the surface, without actually landing. The method involved loosening material and catching it in a horn (see 'Asteroid treasure').

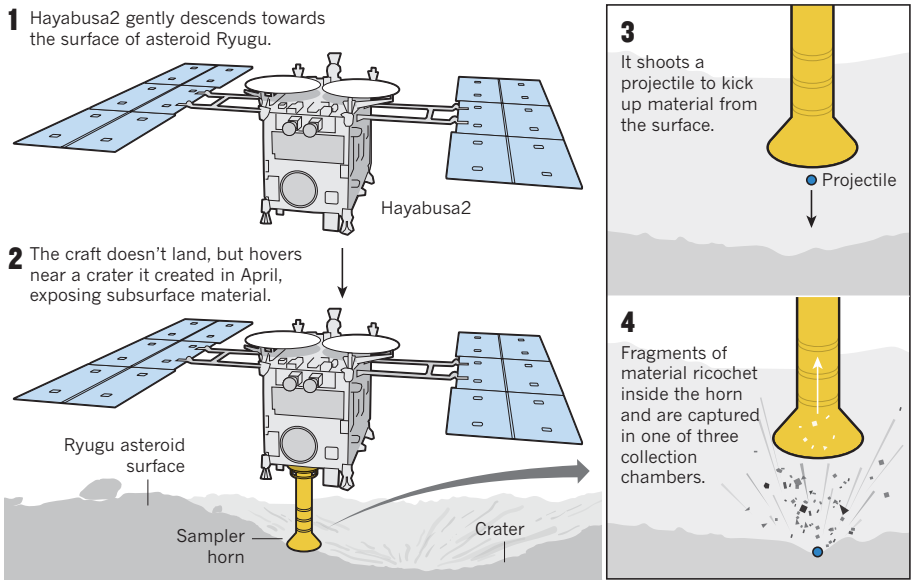
The goal is to bring back a total of around a gram of material. But the team will have to wait until the probe returns to Earth to open the chambers and see what's inside. While Hayabusa2 is in space, mission control has no way of knowing how much material has been collected in each touchdown operation, Yoshikawa says.

Physicists hope that the materials will help to solve asteroid mysteries — for instance, it's not clear why Ryugu is so dark. It is among the least reflective bodies in the Solar System, darker than any known meteorite, and the material exposed at the bottom of the freshly dug crater is darker still. Researchers with JAXA are keen to find out whether the April impact itself made the material darker, or whether the crater's colour is typical of Ryugu's composition and the surface has been lightened by solar radiation.

Ryugu's surface is also strewn with an unusual number of boulders — more per unit surface area than any asteroid explored so far, according to a paper the mission scientists published in May (T. Michikami *et al. Icarus* **331**, 179–191; 2019). This makes the approach and touchdown particularly hazardous for Hayabusa2, especially given that the craft has to operate autonomously owing to its large distance from Earth. ■

ASTEROID TREASURE

With its final major manoeuvre, Japan's spacecraft Hayabusa2 is aiming to become the first probe ever to collect material from beneath the surface of an asteroid.



ARTIFICIAL INTELLIGENCE

AI beats professionals at six-player poker

Triumph in complex variant of game brings bots closer to solving thorny real-world problems.

BY DOUGLAS HEAVEN

Machines have raised the stakes once again. A superhuman poker-playing bot called Pluribus has beaten top human professionals at six-player no-limit Texas hold 'em poker, the most popular variant of the game. It is the first time that an artificial-intelligence (AI) program has beaten elite human players at a game with more than two participants (N. Brown and T. Sandholm *Science* http://doi.org/c766; 2019).

"While going from two to six players might seem incremental, it's actually a big deal," says Julian Togelius at New York University in New York City, who studies games and AI. "The multiplayer aspect is something that is not present at all in other games that are currently studied."

The team behind Pluribus had already built an AI, called Libratus, that had beaten professionals at two-player poker. It built Pluribus by updating Libratus and created a bot that needs much less computing power to play matches. In a 12-day session with more than 10,000 hands, it beat 15 leading human players. "A lot of AI researchers didn't think it was possible to do this" with our techniques, says Noam Brown at Carnegie Mellon University in Pittsburgh, Pennsylvania, and Facebook AI Research in New York City, who developed Pluribus with his Carnegie colleague Tuomas Sandholm.

Other AIs that have mastered human games — such as Libratus and DeepMind's Go-playing bots — have shown that they are unbeatable in two-player zero-sum matches. In these scenarios, there is always one winner and one loser, and game theory offers a well-defined best strategy — use it and you can't lose.

But game theory is less helpful for scenarios involving multiple parties with competing interests and no clear win–lose conditions — which reflect most real-life challenges. By solving multiplayer poker, Pluribus lays the foundation for future AIs to tackle complex problems of this sort, says Brown. He thinks that the success is a step towards applications such as automated negotiations, better fraud detection and self-driving cars.

To tackle six-player poker, Brown ▶

and Sandholm radically overhauled Libratus's search algorithm. Most game-playing AIs search forwards through decision trees for the best move to make in a given situation. Libratus searched to the end of a game before choosing an action.

But the complexity introduced by extra players makes this tactic impractical. Poker requires reasoning with hidden information — players must work out a strategy by considering what cards their opponents might have and what opponents might guess about their hand on the basis of previous betting. But more players makes choosing an action more difficult, because it involves assessing a larger number of possibilities.

The key breakthrough was developing a method that allowed Pluribus to make good choices after looking ahead only a few moves, rather than to the end of the game.

Pluribus teaches itself from scratch using a form of reinforcement learning similar to that used by DeepMind's Go AI, AlphaZero. It starts off playing poker randomly and improves as it works out which actions win more money. After each hand, it looks back at how it played and checks whether it would have made more money with different actions, such as raising rather than sticking to a bet. If the alternatives lead to better outcomes, it will be more likely to choose them in future.

By playing trillions of hands of poker against itself, Pluribus creates a basic strategy that it draws on in matches. At each decision point, it compares the state of the game with its blueprint and searches a few moves ahead to see how the action played out. It then decides whether it can improve on that action.

AI PLAYPEN

Pluribus's success is largely down to its efficiency. When playing, it runs on just two central processing units (CPUs). By contrast, when it first beat leading professionals, DeepMind's original Go bot used nearly 2,000 CPUs; Libratus used 100. When playing against itself, Pluribus plays a hand in around 20 seconds — roughly twice as fast as human professionals.

Games have proved a great way to measure progress in AI because scores can be compared with those of top humans — and bots can objectively be hailed as superhuman if they triumph. But Brown thinks that AIs are outgrowing their playpen. "This was the last remaining challenge in poker," he says.

Togelius thinks there is mileage yet for AI researchers and games. "There's a lot of unexplored territory," he says. Few AIs have mastered more than one game, which requires general ability rather than a niche skill. And there's more than simply playing games, says Togelius. "There's also designing them. A great AI challenge if there ever was one." ■

KOHEI TANAKA/UNIV. TSUKUBA

A bright red rock layer rests above a pocket of orange rock containing fossil egg fragments.

PALAEONTOLOGY

Dinosaurs nested in groups

A site in southeast Mongolia suggests that some dinosaurs guarded their eggs, much like certain modern birds.

BY JONATHAN LAMBERT

An exquisitely preserved dinosaur nesting site discovered in the Gobi Desert shows that some of these prehistoric animals nested in groups and, like birds, protected their eggs.

"Dinosaurs are often portrayed as solitary creatures that nested on their own, buried their eggs and then just went away," says François Therrien, a palaeontologist at the Royal Tyrrell Museum of Palaeontology near Calgary, Canada. He co-authored a study (K. Tanaka *et al. Geology* http://doi.org/c8cc; 2019) published this month in *Geology* describing the find. "But here we show that some dinosaurs were much more gregarious. They came together and established a colony that they likely protected," Therrien says.

The find includes the fossils of 15 nests and more than 50 eggs that are roughly 80 million years old. It provides the clearest evidence yet that group nesting evolved before modern birds split off from the dinosaurs 66 million years ago.

Certain modern birds and crocodiles build nests and lay eggs in a communal area during their breeding seasons. Many palaeontologists think that this 'colonial nesting' first arose in dinosaurs as a way to counter nest predators. But the evidence for this hasn't been solid, says Amy Balanoff, a palaeontologist at Johns Hopkins University in Baltimore, Maryland.

Since the 1980s, palaeontologists have unearthed fossilized eggs or nests that are clustered together. But the surrounding rock often represents several thousand years or more, making it difficult for researchers to tell whether the eggs were laid at the same time, or just in the same place years apart, says Darla Zelenitsky, a palaeontologist at Calgary University in Canada and a study co-author.

The recently described nest site is different. Located in southeast Mongolia, the 286-square-metre formation contains vivid layers of orange and grey rock. Between these bands runs a thin streak of bright red rock that connects 15 clutches of relatively undisturbed eggs. Some of the spherical eggs, about 10–15 centimetres in diameter, had hatched and were partially filled with the red rock.

Flooding from a nearby river blanketed the nesting site under a coating of sediment and probably created the red line, says Therrien.

"Because everything is relatively undisturbed, it likely wasn't a massive flood," he says.

But the streak connects all of the eggs, suggesting that the dinosaurs laid them in a single breeding season. "Geologically, I don't think we could've asked for a better site," says Zelenitsky.

"It's a compelling story," says Balanoff, adding that the researchers back it up with a strong analysis.

The team was also able to identify the type of dinosaur that was probably responsible. The eggs' exterior and interior textures, as well as shell thickness, point to a kind of non-avian theropod, a large group that includes dinosaurs such as velociraptors and *Tyrannosaurus*.

The researchers also estimated that just over half of the nests had at least one successful hatch owing to the number of fragmented eggs. This relatively high rate mirrors the hatching success of modern birds and crocodiles that guard their nests, as opposed to those that abandon or only occasionally check them.

Daniel Barta, a palaeontologist at California State University, Los Angeles, agrees that such a high rate suggests that some dinosaurs tended their nests. But he cautions that eggs that have hatched and those that predators have cracked open can often look similar. ■

ENVIRONMENT

China feels the heat over rogue CFC emissions

The government says it will build a monitoring network to understand what is going on.

BY DAVID CYRANOSKI

When atmospheric models traced a mysterious spike of an ozone-destroying gas to two provinces in China earlier this year, scientists waited to see how the Chinese government, and other nations, would respond to this possible violation of international law.

Now the government is under pressure to act — and has presented a plan to help it track and reduce emissions of the chemical, known as trichlorofluoromethane or CFC-11. Measures include establishing a national monitoring network to track ozone-depleting chemicals, along with heftier penalties for companies caught illegally producing the chemical.

Details of the plan emerged in notes released last month from a May meeting of the Multilateral Fund for the Implementation of the Montreal Protocol in Montreal, Canada.

The document "sets the stage for real progress on this important issue", says David Fahey, director of the Chemical Sciences Division at the US National Oceanic and Atmospheric Administration (NOAA) Earth System Research Laboratory in Boulder, Colorado.

The Chinese environment ministry disputes that there is enough evidence to pin the recently discovered spike in emissions on China, but agrees that more data are needed to understand the problem.

CFC-11 was once a popular refrigerant, and widely used to produce polyurethane foam insulation. But the legally-binding 1987 Montreal Protocol called for its production and trade to be phased out by 2010.

The treaty worked, and global CFC-11 production dropped until 2013, when a surprising slowdown in that decline suggested that there was a new source of emissions.

In a study published in *Nature* in May 2018, researchers traced the spike to east Asia (S. A. Montzka *et al. Nature* **557**, 413–417; 2018). In another published in May 2019, they traced it more precisely, to the provinces of Hebei and Shandong in northeastern China, using data from monitoring stations in Japan and South Korea (M. Rigby *et al. Nature* **569**, 546–550; 2019). Scientists suspect that factories in those provinces might have resumed production of a CFC-11-based foam insulator. Because China has ratified the Montreal Protocol, it is obliged to address any illegal CFC-11 production.

In May, China's environment ministry questioned the conclusions of the most recent study, noting "great uncertainty" in the reported amount and location of emissions. But the statement also said that the study "makes us realize the importance of atmosphere monitoring".

China has released few details about its plan for a national monitoring network. But the documents from the Montreal meeting say routine monitoring will begin within three years in several key cities. If the network uses stations close to Hebei and Shandong, it could pinpoint the source of the mysterious CFC-11 spike, says scientists. That in turn would aid efforts to eliminate the spike, says Stephen Montzka, an atmospheric chemist at the NOAA Earth System Research Laboratory who was part of the team that identified ▶

JG/ALAMY

A source of ozone-destroying gas has been tracked to China, and foam manufacturers are in the firing line.

the spike. Montzka hopes that any data gathered by the national network will be open to the global scientific community.

The Chinese government acknowledges some illegal CFC-11 production: before the Montreal meeting, it reported to the multilateral fund that it had seized 114 tonnes of illegally produced CFC-11 since 2012.

But such amounts could not account for the roughly 7,000 tonnes of CFC-11 that, according to estimates in the 2019 *Nature* paper, is being newly produced each year. (*Nature*'s news team is editorially independent of its journal team.)

Independent scientists say they have confidence in those estimates. "The measurements are of a very high quality," says Claire Reeves, an atmospheric chemist at the University of East Anglia in Norwich, UK. She is leading a team that is building a CFC-11 data set from samples taken in northern Taiwan, which also points to a source of CFC-11 emissions in northeastern China, she says.

A former central-government employee who worked on the regulation of ozone-depleting chemicals, and who asked to remain anonymous because of the sensitivity of the issue, says it is likely there is production of these chemicals that the Chinese government does not know about. Only a few people at local environment and ecology bureaus are assigned to oversee an entire province's monitoring and enforcement efforts on these gases, the source says: "It's not a core task of the ministry."

Attempts by *Nature*'s news team to contact the environment and ecology board for Hebei province were unsuccessful. Shandong province responded to queries about its efforts by forwarding statements from the national environment ministry. The national ministry did not respond directly to the source's comments, but its spokesperson acknowledged that some problems have arisen since the Montreal Protocol was established. "We believe that as part of international collaborative efforts, we can work for the continued success and progress of the Montreal Protocol."

"We can work for the continued success and progress of the Montreal Protocol."

ONE THOUSAND STATIONS

To measure CFC-11, the new network could make use of China's more than 1,000 existing air-quality monitoring stations.

Some are already used to measure CFC-11, but finding the source of the spike that emerged in 2013 would require more frequent sampling over a longer period than is seen for the data from these stations that have appeared in the literature, says Montzka.

China also told the multilateral-fund meeting that it will establish six new laboratories capable of testing for ozone-depleting chemicals in insulation foam, which will speed up the testing of suspect products and help the government to crack down on such activity. Two will be in Beijing and one in Jinan, the capital of Shandong, according to the environment ministry.

Many representatives at the meeting in Montreal, including those from the United States and Japan, the top two contributors to the fund, expressed frustration with China's apparent failure to curb emissions and demanded quick action. Japan even warned that it might withdraw funding if the cause of the CFC-11 emissions is not addressed.

China says that it will report on its progress at the next meeting of the fund in December. ■

CLARIFICATION

The News Feature 'Ancient proteins tell their tales' (*Nature* **570**, 433–436; 2019) did not make clear that the *Stephanorhinus* work cited in reference 6 was led by Enrico Cappellini.

AFTER THE FLAMES

Nearly a year after Brazil's biggest natural-history museum burnt down, researchers are struggling to revive their work and resume their lives.

BY EMILIANO RODRÍGUEZ MEGA

The fire had already chewed up the front half of Brazil's National Museum in Rio de Janeiro by the time zoologist Paulo Buckup drove up. The blaze was surging into the rest of the museum as firefighters stood by looking helpless. "Then I realized why," says Buckup. "They had no water." The two hydrants next to the museum were dry, and engines had to race to a nearby lake to fill up. Buckup knew that the museum's precious collections wouldn't last long.

On the night of 2 September 2018, he and around 40 other scientists, administrators and volunteers checked their fear and broke into the burning building — forming human chains to rescue specimens, computers, freezers and microscopes.

Inside, the museum felt surreal. The only light in the building came from the progressing fire. Buckup rushed through the dark hallways into the inner courtyard, where a lone firefighter tried in vain to extinguish the flames consuming the top floors. The courtyard echoed with loud cracks, and shards of glass rained down, while "a tornado of smoke" erupted out of some interior windows.

Buckup didn't know it yet, but he was witnessing the biggest scientific tragedy ever recorded in Brazil. Soon, hundreds of years of natural history would turn to ash — including much of the nation's most prized records of its past. The fire claimed tens of thousands of the museum's 20 million fossils, animal specimens, mummies and Indigenous artefacts, including recordings of chants in native languages that are no longer spoken. More than two-thirds of the 90 resident researchers lost all of their work and belongings.

Classes for the museum's graduate students resumed a few days later in one of the annex buildings, and admission exams for new students happened on schedule in November. But ten months after the fire, the research community is still struggling to recover. Many scientists have had to shift research topics entirely — often as visitors at institutions in other countries. Buckup and other researchers whose laboratories did not burn have taken in colleagues seeking space for their students and any surviving specimens. And some have begun the painstaking process of restarting collections that had taken two centuries to build. Together, these scientists are trying to revive what once was one of Latin America's largest science collections.

Brazilian researchers are no strangers to this type of misfortune. Fires have consumed

PHOTOGRAPHY BY MARÍA MAGDALENA ARRÉLLAGA FOR *NATURE*

Maria Elizabeth Zucolotto with a massive meteorite that survived the fire.

at least four other science museums and research centres there over the past ten years; and scientists worry that other natural-history collections are also at risk — thanks to a combination of ageing buildings and budget cuts that have put off essential renovations for years.

Many had warned that a similar fate would befall the National Museum, which was established in 1818. "The museum in Rio was a matchbox," says population geneticist Kelly Zamudio at Cornell University in Ithaca, New York, who grew up in São Paulo and typically travels to collections around Brazil for her research. "It was just waiting to happen."

A NIGHT ABLAZE

Buckup, a fish scientist at the Federal University of Rio de Janeiro (UFRJ), was in the middle of writing a grant proposal when an urgent voice message at 7:55 p.m. alerted him to the fire raging at the museum.

He scrolled through his social-media feeds, where people were already posting pictures, and felt strangely relieved by what he saw.

The fire was ripping through the museum's main building, but had not reached the botanical gardens to the south. That area houses a series of buildings, including the herbarium, the library, the archaeology laboratory and the vertebrate department in which Buckup has worked for 25 years. The department's archive of 600,000 fish specimens floating in yellow-tinged alcohol was the only thing keeping him in Rio de Janeiro, an expensive city known for its rampant violence and poor infrastructure. The fish collection would remain untouched by the fire.

Buckup jumped into his SUV and drove. When he got within a kilometre of the museum, he started to see flames. "The sky was full of sparks."

At around 8:40 p.m., he and others decided to kick open a door to enter the rear part of the museum, which had not yet caught fire. They started removing what they could from the teaching department. Another group went to the crustacean laboratory to recover materials. As the blaze slowly ate its way towards them, a technician from the mollusc collection, Claudio Costa, asked Buckup to help him retrieve the precious type specimens — those that serve as the basis for describing new species.

That night, Buckup and Costa carried drawers full of preserved snails, clams and other molluscs to safety. In total, they rescued 760 boxes and vials, including all 664 that contained the type specimens. But they couldn't continue. By around 10 p.m., pieces of burning wood were falling on the volunteers, driving them from the building.

For researchers and students, the museum was more than a workplace, and its destruction has left them reeling. In the competitive world of academia, scientists tend to hide their emotions, says Buckup, but that is no longer true at the museum. Since last September, Buckup has found students and senior colleagues — researchers "that you think you'll never see lowering their defences" — crying. "The tears from all those people are still more disturbing to me than the tragedy itself," says Buckup, who sometimes pauses his story to stop his voice from breaking.

Before the fire, months would go by without him running into researchers from other departments. The building was so massive that they could immerse themselves in their work. Now, nine professors have taken refuge in the ichthyology section. "They lost everything — even their birth certificates," he says.

Palaeontologist Antonio Carlos Fernandes knows the feeling. He spent more than 40 years studying the fossils of corals and other invertebrates, and has continued working as a volunteer researcher at the museum since retiring in 2016. But when a century-old skeleton of a humpback whale fell through the ceiling and into his office during the fire, he lost most of his research materials. Fernandes still finds himself "wanting to believe it was all just a big nightmare". But he has no plans to abandon his work. "Once a researcher, always a researcher," he says.

That's a common sentiment. Members of the entomology department have started to replace their destroyed collections by retrieving some of the specimens that were loaned to other institutions. They have also received generous donations from collectors, and have begun venturing into the Amazon and other regions around Brazil to collect fresh samples. But it will be a challenge to resurrect an inventory that once totalled some 5 million insects — not least because many of the forests that yielded those specimens have since been transformed into farmlands and cities, says museum entomologist Pedro Souza-Dias. "We don't know if we'll find them again."

He has organized six expeditions to the Amazon, Paraná and nature reserves in Rio de Janeiro in the hope of adding more crickets, grasshoppers, mantises and stick insects to the recovering collection. The newly amassed invertebrates are now temporary residents in the already cramped vertebrate department. "We are not in our best conditions right now, but we are fighting," says Souza-Dias. "We don't have another option."

NORTHERN REFUGE

After the fire, Thaynara Pacheco had trouble sleeping. The entomologist was haunted by a burning smell and by the fear that her apartment, like the insect collection, had caught fire. In March, she traded the odour of smoke for the fumes of naphthalene preservative, when she took a fellowship at the Smithsonian National Museum of Natural History (NMNH) in Washington DC.

On a day last March, Pacheco opens a wooden box and reveals hundreds of tiny beetles pinned in place. They belong to the Sericini tribe, which she's trying to catalogue. She brought them all the way from her home state, where they are part of the collection of the Federal University of Mato Grosso in Cuiabá. Others are from Nebraska and Florida. And more will come from California and Canada. Surrounded by trays full of insects, Pacheco removes her glasses to peer into a microscope. Up close, a glossy wing cover adorns the greenish-brown body of a beetle. "That's the beautiful one," she says.

A PhD student from the UFRJ and the National Museum's graduate programme, Pacheco is one of 14 fellows selected to continue their studies at Smithsonian institutions

through a US$250,000 emergency exchange programme. “It gives them a boost, I think,” says NMNH ichthyologist Lynne Parenti, who coordinates the programme.

For Pacheco, that meant completely changing her thesis project. Back in Rio de Janeiro, she had been reviewing the taxonomy of Chelonariidae, or turtle beetles, a little-studied family of almost 300 species. But her notebooks, sketches and more than 1,500 specimens from the National Museum and other institutions disappeared in the fire. “It was a general sense of grief, you know? Like losing someone very dear,” she says.

To continue her new project, Pacheco needs to visit the Zoological Research Museum Alexander Koenig in Bonn, Germany, which houses most of the type specimens for the Sericini tribe. But first she intends to take a step to memorialize the National Museum — by getting a tattoo of the logo of her destroyed lab, or maybe even one of the turtle beetles she used to study.

She’s not the only one. Beatriz Hörmanseder, another NMNH fellow, says that getting inked has helped others to cope with the trauma of the fire. Museu na Pele, or Museum on the Skin, is a project she conceived with a Brazilian tattoo artist, Luís Berbert, to give professors, officials and students a free and indelible memory of their institution. A group of 140 people, including some needle-phobes, have already signed up. “When I started Museu na Pele, everybody was smiling more. They talked about their tattoo, not about their loss,” says Hörmanseder, rolling up her left sleeve. The outline of the museum’s façade drawn in black ink runs across her forearm. Below it is a code, MN 7712-V.

That’s the catalogue number belonging to a 110-million-year-old dwarf crocodile-like reptile unearthed in Brazil’s northeastern state of Ceará. For some two years, Hörmanseder had been painstakingly extricating it from the rock with acid, brushes and dental picks at the National Museum. She suspected the opossum-size creature was an unnamed species — or at least evidence that a previously identified extinct genus had survived 10 million years longer than scientists had thought. “It was a big deal for me,” she says. But the Ceará fossil didn’t make it out of the fire.

She is now completing her studies by describing a fossilized crocodile from Utah. It’s a huge switch in focus in terms of evolution. The Utah fossil is much younger, 35 million years old. By that time, crocodiles lived in rivers, swamps and marshes — unlike their earlier relatives, which were strictly marine or terrestrial.

That is why Hörmanseder, who is set to graduate next year, has been trying to learn about groups of crocodiles she had never studied before. During her four-week Smithsonian fellowship last March, she toured three natural-history museums in the United States in search of ancient specimens she could compare against the Utah crocodile.

“It’s kind of suffocating to have so little time and begin from zero,” she says. But she thinks her endeavour will pay off. Having studied all kinds of prehistoric crocodiles will be of help when she starts her doctorate degree elsewhere, in North America or Germany. “I’ll know everything from all around the world,” she says, and bursts out laughing.

OUT OF THE ASH

Early in the morning after the fire, while an avalanche of reporters interviewed her colleagues, UFRJ astronomer Maria Elizabeth Zucolotto entered the museum’s ruins.

When she walked into the main entrance, she saw nothing but the Bendegó, a colossal 5,360-kilogram iron meteorite discovered in 1784 in northeast Brazil. The space rock had been barely licked by the flames: “A symbol of resistance,” says Zucolotto, curator of the museum’s meteorite collection.

Next door, however, the heat had cremated an exhibition of other prized meteorites. Zucolotto went inside, got down on her knees and blindly ran her hands through the ashes that once were display cases. By touch, she found some smaller meteorites, grabbed them and filled her arms with them. But the firefighters didn’t let her stay long. Plaster was still falling from above.

Those fragments from space were among the first objects to be recovered from the National Museum.

On 18 October, more than a month later, the police allowed Zucolotto to return to her old office. Twisted iron beams and cabinets from the upper floors had crashed into the room. That day, she rescued more meteorites, including one called Angra dos Reis, which is valued at $750,000. It was the second time she had recovered the same rock. The first time was in 1997, after police had seized it from two US dealers who had stolen it from the National Museum and replaced it with a fake.

Zucolotto isn’t the only one sifting through the wreckage. On most days, dozens of trained researchers, armed with brushes and trowels, go through the museum’s debris in search of artefacts. Stationed outside, students sieve the dirt through mesh screens, then clean dusty items and photograph them.

“Incredible as it may seem, we’ve had many happy moments,” says palaeontologist Luciana Carvalho, co-coordinator of the team of nearly 70 people. By the end of June, they had recovered 5,345 objects — pterosaur fossils, ancient human bones, coffee mugs, microscopes, full specimen drawers, Egyptian relics and ceramics from the Amazon.

The effort has taken a physical and emotional toll, says Zucolotto. Some days, she hopes the government will rebuild the museum quickly so she can go back, but she also thinks about retiring and finding a successor to care for the surviving meteorites. In the past few months, she has found joy in adopting a bearded grey dog that had emerged hungry and cheerful around the museum in the days following the fire. “He loves me so much,” she says. “I can’t get rid of him.” Researchers named him Fumaça, or Smoke.

A FIRE FORETOLD

The accident last September is only the latest in a long line of fires that have plagued scientific institutions in Brazil. In May 2010, an inferno destroyed the zoological collection of the Butantan Institute in São Paulo — a research powerhouse responsible for most of the venom antisera and vaccines produced in the country. The centre held the largest repository of snakes that Latin America had ever seen, about 90,000 specimens, representing hundreds of species, some endangered or extinct.

“Most of that is now gone,” says Miguel Trefaut Rodrigues, a herpetologist at the University of São Paulo who worked at Butantan as

Students comb through ash to recover specimens and other objects.

Zoologists Paulo Buckup and Alexandre Pimenta (left) examine mollusc specimens that were saved from the fire, which singed some description labels (right).

a wide-eyed 16-year-old trainee in the 1970s. Although Butantan constructed a new building with fire-prevention systems three years later, the institution never fully recovered. Today, its snake bank houses only 24,000 specimens.

When that institute burnt down, Trefaut Rodrigues and a colleague published a column in a national newspaper warning that something like this could happen again because of the poor state of many of the country's museum buildings. "May this tragedy serve as a lesson," they wrote. They begged the government to take care of other biological facilities, and then listed the ones they thought were most at risk — including the National Museum.

One cause for concern in the future is the Museum of Zoology of the University of São Paulo (MZUSP) and its 10 million specimens. In the early 2000s, when Trefaut Rodrigues was about to step down as the museum's director, he pushed to transfer the collections from the 1940s-era building into a larger, more modern complex. The project was approved, and construction began in 2012, but the economic crisis in 2014 halted work. Today, the new venue exists only as a concrete skeleton.

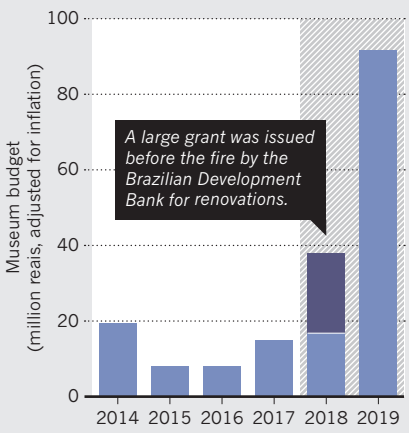
"The university budget now is not enough to finish that thing by any means," says ichthyologist Mario de Pinna, the MZUSP's current director. Still, the museum is taking measures big and small to minimize the danger — from placing heat detectors in all its collections to confiscating coffee machines that represent a risk. "I think we're doing well," says Pinna. "Of course, you know, shit happens. Let's hope it doesn't happen here."

The National Museum had been on a downward spiral for decades, according to museum staff. Critics say that the government ignored many requests over the years to renovate and modernize the facilities. And the financial troubles have only grown. The university's budget, one of the main sources of funding for the museum, has shrunk significantly — from 487 million reais (US$130 million) in 2014, adjusted for inflation, to 361 million reais in 2019. According to the UFRJ, the National Museum was not given enough funds to preserve its collections (see 'Missing money'). "It's not for lack of asking," says Zamudio. "This is the federal government failing science again. They don't want to invest the money. The money, even if it gets appropriated, ends up not reaching the place it should be reaching."

Brazil's Ministry of Education did not respond to *Nature*'s requests to address these criticisms, but did say that it had allocated more than 11 million reais to the National Museum since the fire for response efforts. The ministry also transferred 5 million reais to support reconstruction of the museum.

MISSING MONEY

Critics charge that lack of government support for Brazil's National Museum contributed to the fire risk. Government funding dropped after 2014 — and then surged after the disastrous fire.



But researchers wonder how long the government's commitments will continue. Authorities did not take adequate measures to protect scientific collections in Brazil after the Butantan fire, says Francisco Franco, a biologist and curator of the institute. "As the flames of the fire cooled down, so did the attention of the government," he says. He now fears something similar could happen with the National Museum. "We must not forget."

Buckup never will. One night in March, the museum was at the front of his mind when he joined some of Brazil's most famous personalities, who were being celebrated by the Brazilian newspaper *O Globo* and Rio de Janeiro's industry federation for having "made a difference" in 2018. Buckup went up on stage to accept the honour for his efforts rescuing specimens and equipment last September. But his was not a triumphal speech. "I see no reason to celebrate," he said as he urged the crowd to support the National Museum. "We've lost a part of the past. We can't lose our future."

The forecast isn't good. Even before the museum burnt down, Buckup was losing postdocs and research assistants because of funding cuts. Most moved out of the city; at least one left the country. At the fish collection he curates, even basic maintenance has languished. Buckup says the phones stopped functioning long ago, and Internet access goes down for weeks at a time. What's more, the specimens are preserved at inappropriate temperatures, he says, because the air conditioning remains unfixed.

Another problem worries him, too. Despite numerous requests for maintenance, he says, "the fire protection system is not working". ■

Emiliano Rodríguez Mega *is a science journalist in Mexico City and was an intern with* Nature *in Washington DC.*

SMITA SHARMA FOR *NATURE*

Carl Malamud in front of the data store of 73 million articles that he plans to let scientists text mine.

THE PLAN TO MINE THE WORLD'S RESEARCH PAPERS

A data store in India could open up vast swathes of science for easy computerized analysis.

BY PRIYANKA PULLA

Carl Malamud is on a crusade to liberate information locked up behind paywalls — and his campaigns have scored many victories. He has spent decades publishing copyrighted legal documents, from building codes to court records, and then arguing that such texts represent public-domain law that ought to be available to any citizen online. Sometimes, he has won those arguments in court. Now, the 60-year-old American technologist is turning his sights on a new objective: freeing paywalled scientific literature. And he thinks he has a legal way to do it.

Over the past year, Malamud has — without asking publishers — teamed up with Indian researchers to build a gigantic store of text and images extracted from 73 million journal articles dating from 1847 up to the present day. The cache, which is still being created, will be kept on a 576-terabyte storage facility at Jawaharlal Nehru University (JNU) in New Delhi. "This is not every journal article ever written, but it's a lot," Malamud says. It's comparable to the size of the core collection in the Web of Science database, for instance. Malamud and his JNU collaborator, bioinformatician Andrew Lynn, call their facility the JNU data depot.

No one will be allowed to read or download work from the repository, because that would breach publishers' copyright. Instead, Malamud envisages, researchers could crawl over its text and data with computer software, scanning through the world's scientific literature to pull out insights without actually reading the text.

The unprecedented project is generating much excitement because it could, for the first time, open up vast swathes of the paywalled literature for easy computerized analysis. Dozens of research groups already mine papers to build databases of genes and chemicals, map associations between proteins and diseases, and generate useful scientific hypotheses. But publishers control — and often limit — the speed and scope of such projects, which typically confine themselves to abstracts, not full text. Researchers in India, the United States and the United Kingdom are already making plans to use the JNU store instead. Malamud and Lynn have held workshops at Indian government laboratories and universities to explain the idea. "We bring in professors and explain what we are doing. They get all excited and they say, 'Oh gosh, this is wonderful'," says Malamud.

But the depot's legal status isn't yet clear. Malamud, who contacted several intellectual-property (IP) lawyers before starting work on the depot, hopes to avoid a lawsuit. "Our position is that what we are doing is perfectly legal," he says. For the moment, he is proceeding with caution: the JNU data depot is air-gapped, meaning that no one can access it from the Internet.

Users have to physically visit the facility, and only researchers who want to mine for non-commercial

purposes are currently allowed in. Malamud says his team does plan to allow remote access in the future. "The hope is to do this slowly and deliberately. We are not throwing this open right away," he says.

THE POWER OF DATA MINING

The JNU data store could sweep aside barriers that still deter scientists from using software to analyse research, says Max Häussler, a bioinformatics researcher at the University of California, Santa Cruz (UCSC). "Text mining of academic papers is close to impossible right now," he says — even for someone like him who already has institutional access to paywalled articles.

Since 2009, Häussler and his colleagues have been building the online UCSC Genome Browser, which links DNA sequences in the human genome to parts of research papers that mention the same sequences. To do that, the researchers have contacted more than 40 publishers to ask permission to use software to rifle through research to find mentions of DNA. But 15 publishers have not responded or have denied permission. Häussler is unsure whether he can legally mine papers without permission, so he isn't trying. In the past, he has found his access blocked by publishers who have spotted his software crawling over their sites. "I spend 90% of my time just contacting publishers or writing software to download papers," says Häussler.

Some countries have changed their laws to affirm that researchers on non-commercial projects don't need a copyright-holder's permission to mine whatever they can legally access. The United Kingdom passed such a law in 2014, and the European Union voted through a similar provision this year. That doesn't help academics in poor nations who don't have legal access to papers. And even in the United Kingdom, publishers can legally place 'reasonable' restrictions on the process, such as channelling scientists through publisher-specific interfaces and limiting the speed of electronic searching or bulk downloading to protect servers from overload. Such limits are a big problem, says John McNaught, deputy director of the National Centre for Text Mining at the University of Manchester, UK. "A limit of, say, one article every five seconds, which sounds fast for a human, is painfully slow for a machine. It would take a year to download around six million articles, and five years to download all published articles concerning just biomedicine," he says.

Wealthy pharmaceutical firms often pay extra to negotiate special text-mining access because their work has a commercial purpose, says McNaught. In some cases, publishers allow these firms to download papers in bulk, thus avoiding rate limits, according to a researcher at a pharmaceutical firm who did not want to be identified because they were not authorized to talk to the media. University academics, however, frequently restrict themselves to mining article abstracts from databases such as PubMed. That provides some information, but full texts are much more useful. In 2018, a team led by computational biologist Søren Brunak at the Technical University of Denmark in Lyngby showed that full-text searches throw up many more gene–disease links than do searches of abstracts (D. Westergaard *et al. PLoS Comput. Biol.* **14**, e1005962; 2018).

Scientists must also overcome technical barriers when mining articles. It is hard to extract text from the various layouts that publishers use — something that the JNU team is struggling with right now. Tools to convert PDFs to plain text don't always distinguish clearly between paragraphs, footnotes and images, for instance. Once the JNU team has done it, however, others will be saved the effort. The team is close to completing the first round of extraction from the corpus of 73 million papers, Malamud says — although they will need to check for errors, so he expects the database won't be ready until the end of the year.

> **"OUR POSITION IS THAT WHAT WE ARE DOING IS PERFECTLY LEGAL."**

A WORLD OF POSSIBILITIES

Early enthusiasts are already gearing up to use the JNU depot. One is Gitanjali Yadav, a computational biologist at Delhi's National Institute of Plant Genome Research (NIPGR) and a lecturer at the University of Cambridge, UK. In 2006, Yadav led an effort at NIPGR to build a database of chemicals secreted by plants. Called EssOilDB, this database is today scoured by groups from drug developers to perfumeries looking for leads. Yadav thinks that "Carl's compendium", as she calls it, could give her database a leg-up.

To make EssOilDB, Yadav's team had to trawl PubMed and Google Scholar for relevant papers, extract data from full texts where they could, and manually visit libraries to copy out tables from rare journals for the rest. The depot could fast-forward this work, says Yadav, whose team is currently writing the queries they will use to extract the data.

Srinivasan Ramachandran, a bioinformatics researcher at Delhi's Institute of Genomics and Integrative Biology, is also excited by Malamud's plan. His team runs a database of genes linked to type 2 diabetes; they've been crawling PubMed abstracts to find papers. Now he hopes the depot could widen his mining net.

And at the Massachusetts Institute of Technology (MIT) in Cambridge, a team called the Knowledge Futures Group says it wants to mine the depot to map how academic publishing has evolved over time. The group hopes to forecast emerging areas of research and identify alternatives to conventional metrics for measuring research impact, says team member James Weis, a doctoral student at MIT Media Lab.

A CAREER UNLOCKING COPYRIGHT

Malamud only recently had the idea of extending his activism to academic publishing. The founder of a non-profit corporation called Public Resource, based in Sebastopol, California, Malamud has focused on buying up government-owned legal works and publishing them. These include, for instance, the state of Georgia's annotated legal code, European toy-safety standards and more than 19,000 Indian standards for everything from buildings and pesticides to surgical equipment.

Because these documents are often a source of revenue for government agencies, some of them have sued Malamud, who has argued back that documents which have the force of the law cannot be locked behind copyright. In the Georgia case, a US appeals court cleared him of infringement charges in 2018, but the state appealed, and the case is with the US Supreme Court. Meanwhile, a German court ruled in 2017 that the publication of toy standards by Public Resource, including a standard on baby dummies (pacifiers), was illegal.

But Malamud has enjoyed victories, too. In 2013, he filed a lawsuit in a US federal court asking the Internal Revenue Service (IRS) to publish the forms it collected from tax-exempt non-profit organizations — data that could help to hold these organizations to account. Here, the court ruled in Malamud's favour, prompting the IRS to release the financial information of thousands of non-profit organizations in a machine-readable format.

In early 2017, aided by the Arcadia Fund, a London-based charity that promotes open access, Malamud turned his attention to research articles. Under US law, works by US federal government employees cannot be copyrighted, and Public Resource says it has found hundreds of thousands of academic articles that are US government works and seem to defy this rule. Malamud has called for such articles to be freed from copyright assertions, but it's not clear whether that would hold up in court. He has posted his preliminary results online, but has put further campaigning on hold, because the project prompted him to take on a wider mission: democratizing access to all scientific literature.

OPPORTUNITY IN INDIA

A trigger for this mission came from a landmark Delhi High Court judgment in 2016. The case revolved around Rameshwari Photocopy Services, a shop on the campus of the University of Delhi. For years, the business had been preparing course packs for students by photocopying pages from expensive

textbooks. With prices ranging between 500 and 19,000 rupees (US$7–277), these textbooks were out of reach for many students.

In 2012, Oxford University Press, Cambridge University Press and Taylor and Francis filed a lawsuit against the university, demanding that it buy a license to reproduce a portion of each text. But the Delhi High Court dismissed the suit. In its judgment, the court cited section 52 of India's 1957 Copyright Act, which allows the reproduction of copyrighted works for education. Another provision in the same section allows reproduction for research purposes.

Malamud has a long association with India: he first travelled there as a tourist in the 1980s, and he wrote one of his first books, on database design, on a houseboat in Srinagar. And around the same time that he heard about the Rameshwari judgment, he had come into possession (he won't say how) of eight hard drives containing millions of journal articles from Sci-Hub, the pirate website that distributes paywalled papers for anyone to read. Sci-Hub itself has lost two lawsuits against publishers in US courts over its copyright infringements, but despite those judgments, some of its domains are still working today.

Malamud began to wonder whether he could legally use the Sci-Hub drives to benefit Indian students. In a 2018 book about his work called *Code Swaraj*, co-authored with Indian tech entrepreneur Sam Pitroda, Malamud writes that he imagined showing up on Indian campuses in the equivalent of an American taco truck, ready to serve the articles up to those who wanted them.

Ultimately, he zeroed in on the idea of the JNU text-mining depot instead. (Malamud has also helped to set up another mining facility with 250 terabytes of data at the Indian Institute of Technology Delhi, which isn't in use yet.) But he is cagey about where the depot's articles come from. Asked directly whether some of the text-mining depot's articles come from Sci-Hub, he said he wouldn't comment, and named only sources that provide free-to-download versions of papers (such as PubMed Central and the 'Unpaywall' tool). But he does say that he does not have contracts with publishers to access the journals in the depot.

IS IT LEGAL?

Malamud says that where he got the articles from shouldn't matter anyway. The data mining, he says, is non-consumptive: a technical term meaning that researchers don't read or display large portions of the works they are analysing. "You cannot punch in a DOI [article identifier] and pull out the article," he says. Malamud argues that it is legally permissible to do such mining on copyrighted content in countries such as the United States. In 2015, for instance, a US court cleared Google Books of copyright infringement charges after it did something similar to the JNU depot: scanning thousands of copyrighted books without buying the rights to do so, and displaying snippets from these books as part of its search service, but not allowing them to be downloaded or read in their entirety by a human.

SAJJAD HUSSAIN/AFP/GETTY

Rameshwari Photocopy Services in New Delhi was taken to court for copying parts of textbooks, and won.

The Google Books case was a test of non-consumptive data mining, says Joseph Gratz, an IP lawyer at the law firm Durie Tangri in San Francisco, California, who represented Google in the case and has previously represented Public Resource. Even though Google was displaying snippets, the court ruled that the text was too limited to amount to infringement. Google was scanning authorized copies of books (from libraries in many cases), even though it did not ask permission. Copyright holders might argue that if Sci-Hub or other unauthorized sources supplied the JNU depot, the situation would be different from the Google Books case, Gratz says. But a case involving unauthorized sources has never been argued in American courts, making it hard to predict the outcome. "There are good reasons why the source shouldn't matter, but there may be arguments that it should," says Gratz.

The question of the facility's legality in the United States might not even be relevant, because international researchers would be getting results from a depot that sits in India, even if they are accessing it remotely. So Indian law is likely to apply to the question of whether it is legal to create the corpus, says Michael W. Carroll, a professor at the American University's Washington College of Law in Washington DC.

Here, India's copyright laws might help Malamud — another reason why the facility is in New Delhi. The research exemption in section 52 means that the JNU data depot's actions would be considered fair use of copyrighted material under Indian law, argues Arul George Scaria, an assistant professor at Delhi's National Law University. Not everyone agrees with this interpretation, however. Section 52 allows researchers to photocopy a journal article for personal use, but doesn't necessarily allow the blanket reproduction of journals as the JNU depot has done, says T. Prashant Reddy, a legal researcher at the Vidhi Centre for Legal Policy in New Delhi. That entire articles aren't shared with users does help, but the mass reproduction of text used to create the database puts the facility in "a legal grey zone", Reddy says.

RISKY BUSINESS

When *Nature* contacted 15 publishers about the JNU data depot, the six who responded said that this was the first time they had heard of the project, and that they couldn't comment on its legality without further information. But all six — Elsevier, BMJ, the American Chemical Society, Springer Nature, the American Association for the Advancement of Sciences and the US National Academy of Sciences — stated that researchers looking to mine their papers needed their authorization. (Springer Nature publishes this journal; *Nature*'s news team is editorially independent of its publisher.)

Malamud acknowledges that there is some risk in what he is doing. But he argues that it is "morally crucial" to do it, especially in India. Indian universities and government labs spend heavily on journal subscriptions, he says, and still don't have all the publications they need. Data released by Sci-Hub indicate that Indians are among the world's biggest users of their website, suggesting that university licences don't go far enough. Although open-access movements in Europe and the United States are valuable, India needs to lead the way in liberating access to scientific knowledge, Malamud says. "I don't think we can wait for Europe and the United States to solve that problem because the need is so pressing here." ■

Priyanka Pulla *is a freelance journalist based in Bengaluru, India.*

COMMENT

J. A. ROGERS/NORTHWESTERN UNIV.

Sensors soft enough to be used on a premature baby's skin can monitor vital signs in the neonatal intensive care unit.

Skin sensors are the future of health care

Thin, flexible, wireless monitoring systems could make medicine more predictive and personalized, argue **Shuai Xu**, **Arun Jayaraman** and **John A. Rogers**.

Thin, soft electronic systems that stick onto skin are beginning to transform health care. Millions of early versions[1] of sensors, computers and transmitters woven into flexible films, patches, bandages or tattoos are being deployed in dozens of trials in neurology applications alone[2]; and their numbers growing rapidly. Within a decade, many people will wear such sensors all the time. The data they collect will be fed into machine-learning algorithms to monitor vital signs, spot abnormalities and track treatments.

Medical problems will be revealed earlier. Doctors will monitor their patients' recovery remotely while the patient is at home, and intervene if their condition deteriorates. Epidemic spikes will be flagged quickly, allowing authorities to mobilize resources, identify vulnerable populations and monitor the safety and efficacy of drugs issued. All of this will make health care more predictive, safe and efficient.

Where are we now? The first generation of biointegrated sensors can track biophysical signals, such as cardiac rhythms, breathing, temperature and motion[3]. More advanced systems are emerging that can track certain biomarkers (such as glucose) as well as actions such as swallowing and speech.

Small companies are commercializing soft biosensor systems that measure clinical data continuously. These include Vital Connect in San Jose, California; iRhythm in San Francisco, California; MC10 in Lexington, Massachusetts; and Sibel Health in Evanston, Illinois. For example, ▶

iRhythm's single-use Zio patch monitors electrical pulses from the heart for 14 days, and is more effective than intermittent hospital check-ups at detecting abnormal rhythms[4]. But it is bulky and temporary, and the data must be downloaded after use, rather than transmitted in real time.

More advanced sensors from our labs are undergoing clinical trials in Chicago, Illinois[5]. These include even smaller sensor networks for heart rate, respiration and temperature. They can transmit data wirelessly, and are soft enough to place on the chests of premature babies without damaging their fragile skin[6]. There is no need for nurses, doctors or parents to disconnect a forest of wires when they want to pick up a baby. Similar systems might sense pressure and temperature in people who have had limbs amputated, at the interface between a limb socket and prosthesis.

Many challenges must be overcome to make wearable sensors fit for widespread use. Innovations in materials, devices and circuit designs must make soft biosensors even smaller, thinner, lighter and less power-hungry. The accuracy, precision and range of measurements must improve. And regulation, costs, usability and data security require attention.

Here, we outline the priorities for action.

TO-DO LIST

Biomarkers. All the flexible sensor systems approved by the US Food and Drug Administration (FDA) so far collect biophysical signals. Biochemical signatures, such as glucose or hormone levels, are hard to glean without puncturing the skin with needles.

Some emerging devices collect fluid by inserting a filament into the skin. And detecting chemicals in sweat is a promising alternative[7]. Sweat contains many indicators relating to cell health and organ function (such as electrolytes), the immune system (cytokines) and drug interactions (metabolites). Sweat sensors are being developed that capture chloride, glucose, lactate, urea, creatinine, alcohol, pH and even heavy metals. Quantifying protein and hormone levels in sweat would increase these sensors' applicability further.

Still, sensors need to be able to collect and analyse sweat without it becoming contaminated or degrading, and they will also require new chemical tests and types of assay.

Tools. Imaging and spectroscopy capabilities would allow for real-time assessments of the body. Examples are optical coherence tomography, confocal microscopy, Raman spectroscopy and two-photon excitation microscopy. If such systems could be miniaturized, they could diagnose skin tumours without the need for a biopsy sample or surgery. They are currently still expensive, bulky and wired.

Therapies. Interfaces that create skin sensations, such as vibrations, might enhance rehabilitation, notably with speech and motion therapies. Drugs could be delivered through skin patches, as they are already for motion sickness (scopolamine), pain (fentanyl), contraception (norelgestromin and ethinylestradiol) and high blood pressure (clonidine). The release could be triggered electrically, acoustically or thermally, for example, by applying heat to a polymer pocket. Sensors could also deliver electrical or thermal stimulation to treat neurological disorders or modulate pain.

Implants. Soft sensor systems could be used inside the body. A thin, flexible implant might be wrapped around the heart or spine to monitor and stimulate it. Demonstration versions of thin, flexible technologies that track the electrical activity of the brain have been tested in mice, cows and non-human primates. Practical challenges include developing biocompatible materials and manufacturing ultra-thin layers that protect the electronics for years or decades. Some patches might melt away harmlessly after they have done their job, just as a wound heals.

Materials and design. There is work to be done to make devices less perceptible to wearers. Today's patches typically include ultra-thin silicon electronics in a matrix of silicone elastomers. In future, organic polymers could be used to make biosensors that repair themselves. And the soft materials will have functions of their own, perhaps being antimicrobial or able to change colour if a biochemical is detected. Power could be harvested from body motions or changes in heat or blood flow rather than from batteries[8].

"Health-care funders should champion biointegrated sensor systems because they can potentially improve the quality of care and lower costs."

Data. Combinations of sensors need to be designed to suit certain conditions. For example, for Parkinson's disease, a single sensor on the hand is enough to detect tremors[9]. But in people who have had a stroke, characterizing how hard their foot hits the ground when walking, how strongly they swallow or how soundly they sleep would require additional sensors and data outputs — from accelerometers, gyroscopes, microfluidic sensors, and electrocardiographs and electromyographs (which measure electrical activity in the heart and muscles, respectively). To improve data quality, these sensors should be sited on the best places on the body to collect information — for example, electrocardiogram signals should be recorded on the chest, not the wrist. Gait is better assessed with sensors on the ankles. Noise will need to be filtered out, and decisions will need to be made about whether it is better to stream all of the data to the cloud or process some of them on the chip and transmit only key parameters or insights extracted from the base data, in the form of warnings or notices.

Interpretation. Digital dashboards need to be developed to allow physicians and patients to track outputs, log changes and make clinical decisions. Machine-learning models need improvement, for example to predict how long it will be until a patient is discharged from hospital or is able to walk or feed themselves safely without assistance. Long-term monitoring in the community would help physicians to assess the evolution of stroke recovery, Parkinson's disease and other disorders.

Behaviour. More needs to be learnt about how patients use biosensors in their everyday lives. If people are to wear the devices for weeks or months, the patches will need to look acceptable, and ideally attractive. They should be comfortable and maintain good contact with the skin during washing or exercising. Although some sensors are now small enough fit on a fingernail and thin enough not to show through clothing, they will need to become yet smaller and thinner.

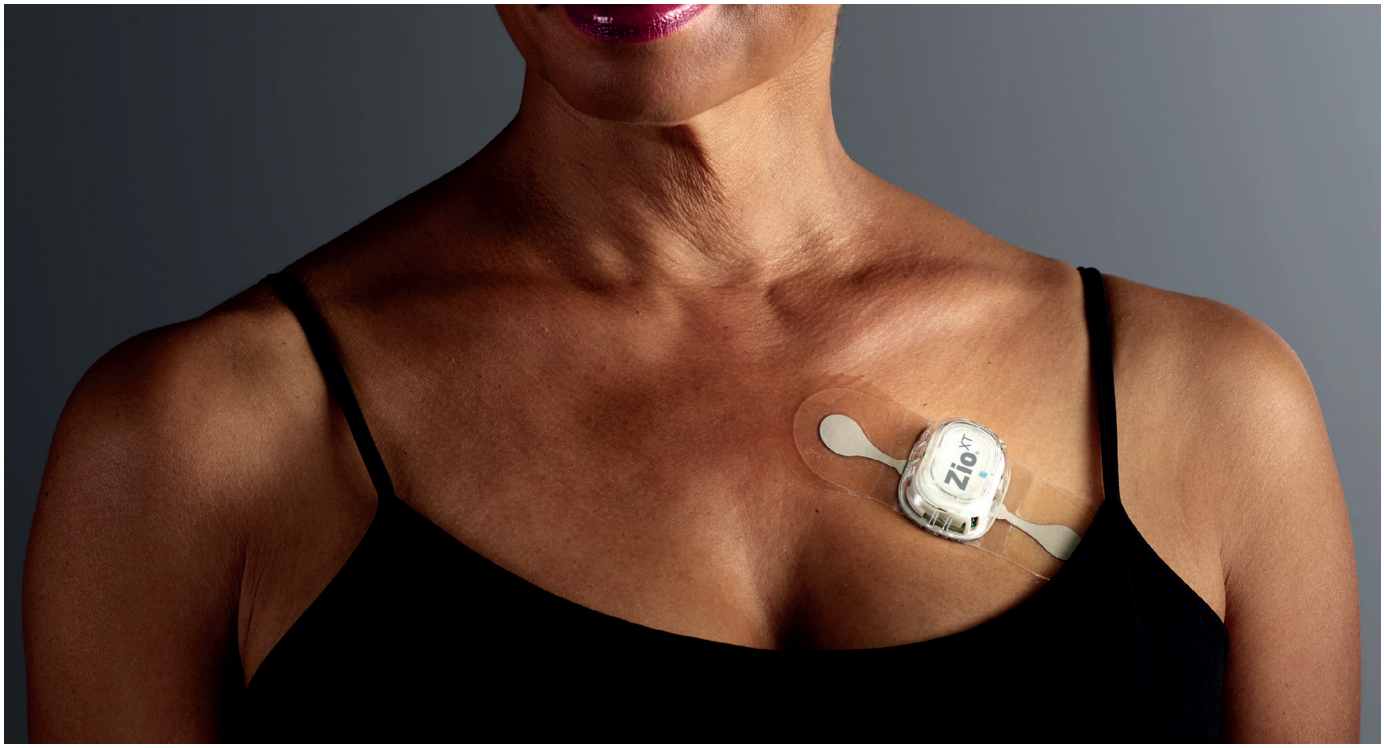
CLINICAL PRACTICE

Bringing these technologies to patients will take action on three more fronts: validation, regulation and data protection.

To speed up their entry into the clinic, soft biosensors must target unmet medical needs, such as mental-health monitoring in the home[10]. Changes in vital signs and in neuroendocrine, neurotropic and inflammatory biomarkers could yield insights that are unavailable to clinicians today. Signs of social isolation and loneliness might prompt a visit from a carer or a call from a loved one.

Wireless health monitoring could also revolutionize health care in countries where infrastructure is lacking. We will trial our biosensors in maternity clinics in several African countries, including Zambia, Kenya and South Africa, later this year, in partnership with the non-profit organizations the Bill & Melinda Gates Foundation and Save the Children. The patches will track physiological data such as physical activity, blood pressure and respiratory rate in women and their babies during pregnancy, warning of complications such as fetal hypoxia or an impending haemorrhage.

IRHYTHM TECHNOLOGIES



The single-use Zio patch from iRhythm in San Francisco, California, can monitor heart rate continuously for two weeks to detect irregularities.

Regulatory approval is crucial, and challenging to obtain. Hardware is largely covered by existing frameworks; algorithms are not. But there are encouraging signs that software applications can be regulated. In the past few years, the FDA has approved machine-learning technology for the diagnosis of diabetic retinopathy, the first pill with an embedded sensor (Abilify MyCite) and an app to treat opioid-use disorder (reSET-O). The FDA's pre-certification programme allows medical software from certain trusted developers to be deployed before formal evaluation.

Regulations must adapt quickly, as the boundaries between devices, data, software and therapeutics continue to blur. Special attention should be paid to clinical areas of highest need and minimal risk — there are some such within rare diseases, paediatrics, women's health and gerontology.

Data security must be a top priority, particularly for patient information. The US Health Insurance Portability and Accountability Act established guidelines for the confidential handling of patient information in 1996. But this was well before the explosion in mobile devices and wearable sensors. New frameworks are needed. Patients must own their own data. And great care must be taken to ensure that companies do not exploit medical data for commercial gain without approval, or drive a division between those who can and cannot access this technology.

Given the poor track record of private companies in protecting consumer privacy, leadership at both the national and international level is needed. Policies must prevent employers and insurers from discriminating against people with particular data profiles, much as the US Genetic Information Nondiscrimination Act of 2008 protects workers. Deviations should be met with serious financial and legal punishments[11].

It remains to be seen how these sensor systems will be paid for, and how doctors will be reimbursed for interpreting and acting on the data. Still, health-care funders should champion biointegrated sensor systems because they can potentially improve the quality of care and lower costs. This fits with the move towards value-based care in the United States, where health-insurance companies and government plans such as Medicare are selecting treatments on the basis of efficacy rather than simply reimbursing services.

ROAD AHEAD

Technical progress will require close collaborations between materials and device engineers, data scientists and medical professionals. Users and carers need to be more closely involved.

Interdisciplinary funding from government sources, corporate investments and charitable foundations will be essential for collecting proof-of-concept data before devices can be commercialized. For example, the Michael J. Fox Foundation in New York City has grant programmes focused on wearable technologies in global health.

Companies need to improve manufacturing processes for devices that combine hard skeletal components and soft tissue-like materials. Yields and throughputs need to be improved to assure quality and lower costs. Automated tools are needed to design the layout and topology of circuits and mechanical components.

The effort will be worth it: bio-integrated sensors have the potential to transform nearly every aspect of medicine. ■

Shuai Xu *is the medical director at the Center for Bio-Integrated Electronics, Northwestern University, Evanston, Illinois, USA.* **Arun Jayaraman** *is the director of the Max Nader Center for Rehabilitation Technologies and Outcomes Research, Shirley Ryan Ability Lab, Chicago, Illinois, USA.* **John A. Rogers** *is the Simpson-Querrey Professor of Materials Science and Engineering at the Center for Bio-Integrated Electronics, Northwestern University, Evanston, Illinois, USA.*
e-mail: jrogers@northwestern.edu

1. Topol, E. *Nature Med.* **25**, 44–56 (2019).
2. Johansson, D., Malmgren, K. & Alt Murphy, M. *J. Neurol.* **265**, 1740–1752 (2018).
3. Ray, T. R. *et al. Chem. Rev.* **119**, 5461–5533 (2019).
4. Steinhubl, S. R. *et al. J. Am. Med. Assoc.* **320**, 146–155 (2018).
5. Rogers, J., Malliaras, G. & Someya, T. *Sci. Adv.* **4**, eaav1889 (2018).
6. Chung, H. U. *et al. Science* **363**, eaau0780 (2019).
7. Heikenfeld, J. *et al. Nature Biotechnol.* **37**, 407–419 (2019).
8. Wu, W. & Haick, H. *Adv. Mater.* **30**, e1705024 (2018).
9. Lonini, L. *et al. NPJ Digit. Med.* **1**, 64 (2018).
10. Mohr, D. C., Zhang, M. & Schueller, S. M. *Annu. Rev. Clin. Psychol.* **13**, 23–47 (2017).
11. Wilbanks, J. T. & Topol, E. J. *Nature* **535**, 345–348 (2016).

S.X. and J.A.R. declare competing financial interests: see go.nature.com/2lqv2tj for details.

RENAISSANCE SCIENCE

Leonardo's laboratory: studies in flow

On the 500th anniversary of the Renaissance icon's death, **Martin Kemp** looks anew at his innovative experimental models for the motion of water and blood.

Leonardo da Vinci (1452–1519) was a man before his time: we know the cliché. However, in crucial respects he was very much a man of his time. His versatility was foreshadowed by the great artist-engineers of the Italian Renaissance. Notably, Filippo Brunelleschi, inventor and architect of the massive dome of Florence Cathedral, formulated the science of linear perspective for painters in the early years of the fifteenth century. In his work on physical sciences, Leonardo was heir to medieval theories of statics and dynamics; Isaac Newton was still far away. Leonardo's anatomical researches merged medieval physiology with the functional and morphological analyses of the classical physician Galen.

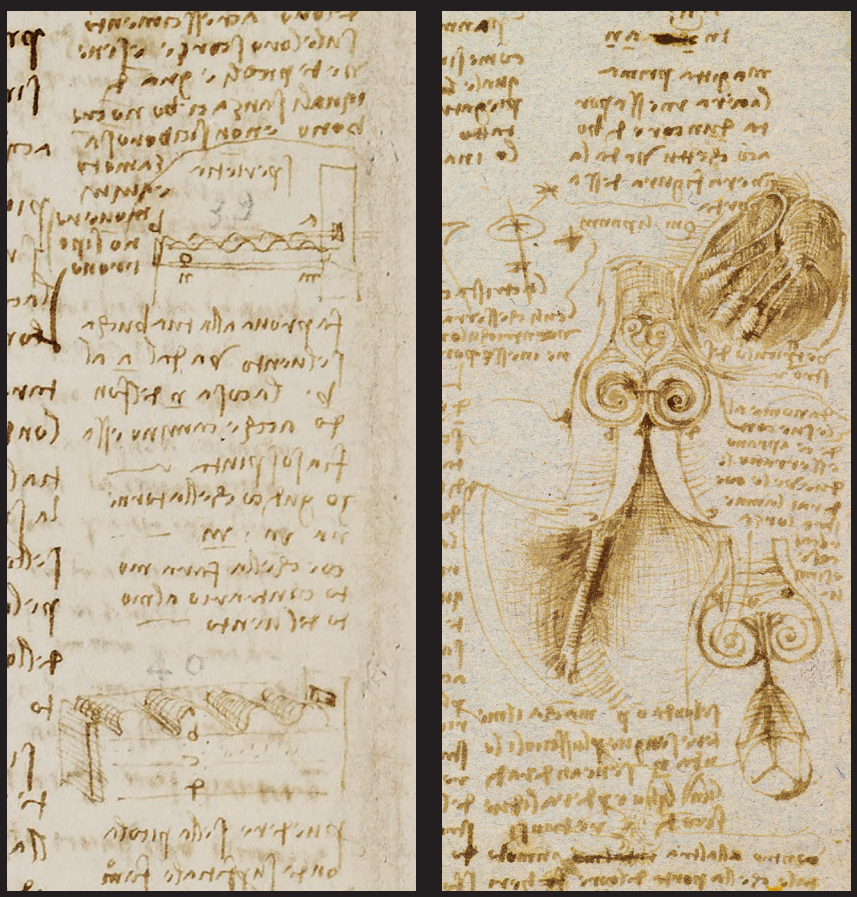
It is better to say that Leonardo's innovations show that he realized more than his predecessors or contemporaries managed to do in the scientific context of his period.

The most conspicuous of Leonardo's innovations is his perfection or invention of almost all the illustrative techniques known before the X-ray. In his notebooks, he depicted subjects using perspective; showed solid forms modelled systematically in light and shade; sectioned them to reveal their inner structures; used transparency to show underlying features; portrayed 'exploded' views of body and machine parts to disclose their forms and articulations; invented diagrammatic representations to disclose the functions of bodily and mechanical systems; and drew thought experiments to explore how things worked. He deployed these techniques widely across his scientific and technical endeavours.

FLUID MOTION

Among the vast range of phenomena that Leonardo explored and depicted was the behaviour of liquids. Now, in *Leonardo da Vinci's Codex Leicester* — a new four-volume edition of the 72-page scientific notebook, composed after 1508 — science historian Domenico Laurenza and I show how his revolutionary ideas on fluid dynamics operated in the specific context of the ancient history of the 'body of the world'.

The outer pages of the codex deal in part with Leonardo's theories on the passage of light from the Sun to Earth and the Moon, involving reflections from seas real or putative. Most of the notebook, however, is devoted to the study of water in motion, in seas, rivers and canals in the form of *vene*



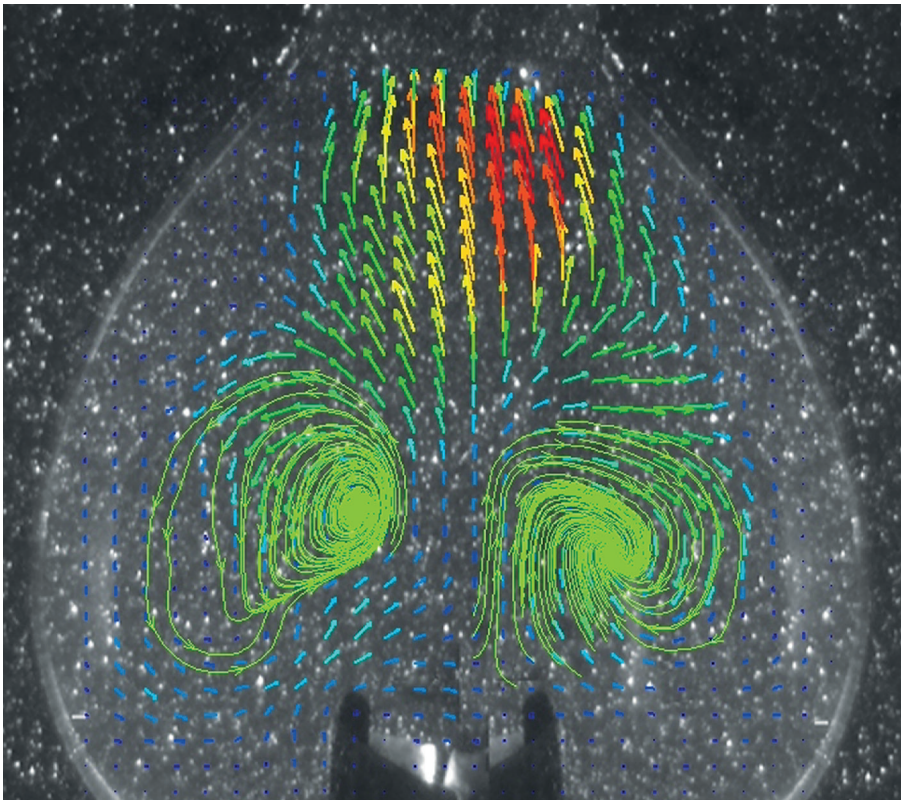
Left: Leonardo da Vinci's sketches of experimental tanks for studying waves and currents in water. Right: Sketches exploring Leonardo's theories about vortices in heart valves.

LEFT: COLLECTION OF BILL GATES, SEATTLE/PHOTO: BILL GATES; RIGHT: LEONARDO DA VINCI, *STUDIES OF BLOOD IN AN AORTIC VALVE AND A GLASS MODEL OF THE NECK OF AN AORTA*, WINDSOR CASTLE, ROYAL COLLECTION TRUST/© HER MAJESTY QUEEN ELIZABETH II 2019

Studies of Turbulent Water **shows Leonardo's understanding of how vortices move.**

LEONARDO DA VINCI, *STUDIES OF TURBULENT WATER*, ROYAL COLLECTION TRUST/© HER MAJESTY QUEEN ELIZABETH II 2019

MORY GHARIB AND ARASH KHERADVAR



A model showing vortices that close a heart's aortic valve, built by Morteza Gharib to Leonardo's design.

d'aqua (veins of water), on Earth's surface and underground. The principle underlying Leonardo's thinking is that of the micro- and macrocosm: he saw the human body as a 'lesser world', mirroring the forms and functions of the wider world.

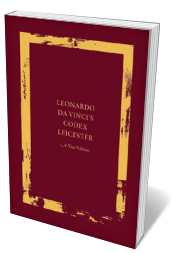
Whereas earlier authorities, such as the second-century Roman astronomer and geographer Ptolemy, had seen Earth as undergoing relatively local changes, Leonardo saw it as having an ancient life-story of vast transformations. In his geological theories, sections of Earth's crust collapsed, violently transforming the relationship between earth and water. As centres of gravity shifted, portions of crust extruded to form lands and mountains.

As Laurenza and I reveal, among the most original aspects of his investigation are what may legitimately be called 'laboratory experiments' using ingenious physical models. Leonardo's extended analyses of the behaviour of water combine mathematical theories of motion with acute observation. The experiments documented in the codex were designed to show how waves are generated by wind, and how currents and vortices perform their complex arabesques under the surface. He sketched an experimental tank in two small drawings in the margins of folio 9v (r denotes the front and v the back of the pages, or folios, of Leonardo's notebooks). The upper drawing is labelled *experientia* ('experience' or 'experiment'). An accompanying note reminds Leonardo "to get a terracotta trough, with a large and flat bottom, 2 braccia [116 centimetres] long and half a braccio [29 cm] wide; have it made here, by the ceramicist". The sides of the tank, he wrote, should be made of glass; he would infuse the water with seeds of panic grass (*Panicum* spp.) that he could use to track the vortices in action.

VALVES AND VORTICES

One of the tasks he set himself was to observe what happens to a mobile object at the bottom of the tank when a current of air is blown through a rectangular opening at one end, just above the water's surface. He decided that the object would move in an opposite direction from the wind.

The experimental set-ups reveal that Leonardo's magnificent studies of turbulence, kept in the Royal Library at Windsor Castle, UK, stem not from observation in nature but from experiments on vortex motions. Concentrating in turn on various aspects of compound action in the motion of water and submerged air, he completed a great synthesis.



Leonardo da Vinci's Codex Leicester: A New Edition (Volume 1)
DOMENICO LAURENZA & MARTIN KEMP
Oxford University Press (2019)

There is more. On folio 15r Leonardo notes that at the junction of two rivers, "what happens with the beds … can be demonstrated with a simple experiment using sand". This laboratory modelling of water interacting with its sandy bed is taken to a higher level of complexity in the Codex Atlanticus, a series of notebooks held in the Ambrosia Library in Milan, Italy. Here, on folio 227v, he proposes to make a proportionally scaled experimental model (*la sperienza nelle minute dimostrazione*) of the gulfs and seas in the Mediterranean, with the major rivers, to test his reconstruction of ancient geological processes. He speculated that the Strait of Gibraltar would widen over time, allowing the Mediterranean to become a mighty river, an extension of the Nile.

Leonardo's studies of the motion of blood in the human body involved similar experimental modelling. He was particularly interested in the heart's passive, three-cusp aortic valve, which he realized must be operated by the motion of blood. To demonstrate his theory that vortices curl back to fill the cusps in the flask-shaped constriction at the aorta's neck, he proposed to make a ceramic mould in the shape of the neck, in which he could blow a glass vessel. He would thus be able to witness (again, courtesy of grass seed) the motion of water and the action of cusps, as a 'proof' of how blood behaves (illustration). Leonardo's model has been constructed by fluid-dynamics specialist Morteza Gharib, who used modern imaging techniques to demonstrate the existence of the revolving vortices that Leonardo interpreted as closing the valve.

Such modelling was hugely original for the early Renaissance. Thus, although Leonardo resided in his own time, he did have some footholds in the future.

It is often claimed that because Leonardo's science was unpublished and uncirculated for centuries, it had no impact on scientific developments. In our edition of the Codex Leicester, Laurenza reveals that the work was in fact available in the circles of those who reformed geology between the seventeenth and nineteenth centuries, such as the volcanologist William Hamilton, British ambassador to Naples from 1764 to 1800. Handwritten copies of the codex enjoyed wide circulation in key places for modern geology: London, Rome, Florence, Naples, Paris and Weimar.

Leonardo always surprises those who study him. ■

Martin Kemp *is emeritus research professor in the history of art at the University of Oxford, UK. He has written and broadcast extensively on imagery in art and science from the Renaissance to today, and written many books on Leonardo da Vinci. His latest, with Domenico Laurenza, is* Leonardo da Vinci's Codex Leicester: A New Edition.
e-mail: martin.kemp@trinity.ox.ac.uk

HULTON ARCHIVE/GETTY

Ford cars under construction in Michigan in 1927.

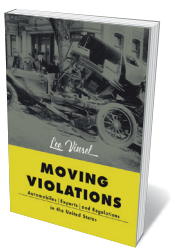
AUTOMOBILE TECHNOLOGY

Driven to distraction

Jack Stilgoe applauds a timely and alarming chronicle of the history of US road safety.

The story told by countless advertisements is that cars mean freedom — open roads, high speeds and boundless possibilities. The paradox is that, when we drive, we are perhaps more constrained than in any other area of everyday life. The German philosopher Max Horkheimer wrote in 1947, "It is as if the innumerable laws, regulations and directions with which we must comply were driving the car, not we."

The shape of a car, the behaviour of a driver, the design of a traffic light and ultimately the configuration of a city are products not just of industrial ingenuity, but of an unfinished tussle between companies and regulators. Many of the rules and standards embodied in car design aim to keep us safe. Yet being in or around cars is among the most dangerous things we do. Some 1.35 million people die on the world's roads every year. Developers of autonomous vehicles think this problem can be solved through artificial intelligence, but their technologies bring new safety concerns. Now, in *Moving Violations*, technology historian Lee Vinsel explains how the car became a fact of life in the United States, and how US regulators shaped this essential component of the American dream in an attempt to mitigate its extraordinary dangers.



Moving Violations: Automobiles, Experts, and Regulations in the United States
LEE VINSEL
John Hopkins University Press (2019)

In 1900, only 8,000 cars were registered in the United States. Handmade by businesses such as the Duryea Motor Wagon Company, these automobiles were playthings of the rich. And risk was part of the appeal. It quickly became clear that a privileged few were creating a public menace. A chaotic aftermarket of safety widgets offered drivers protection in what were then low-speed crashes; but more than 1,000 people were killed by motor vehicles in the United States in 1909. A year before that, the arrival of the Ford Model T made cars affordable for the middle classes, and by 1913 more than one million cars were registered. Safety became a major concern.

For Henry Ford's mass production to work, it required standardization. Screws had to fit regardless of which factory they came from. And this was part of a bigger story — the standardization of streets, traffic signals, drivers and pedestrians. Road safety started to be acknowledged as a problem but, in their battle with government, carmakers found it easy to offload responsibility onto others. John O'Brien, the inspector in charge of traffic for the New York City Police Department in the 1920s, explained that the priority was to "educate the pedestrians". However, through the concerted action of the insurance industry, regulators and researchers in organizations such as the Society of Automotive Engineers (founded in 1905), carmakers were persuaded that safer cars would be good for their bottom line. From the 1930s to the 1960s, the US road death toll climbed, but more slowly than the number of kilometres travelled. (It peaked at more than 50,000 deaths a year in the mid-1970s, then slowly declined.)

RESEARCH ROAD MAP

As Vinsel shows, a history of science is woven through this regulatory tale. Researchers funded by public bodies such as what is now the US National Academies of Sciences, Engineering, and Medicine were confident that they could control risk through scientific knowledge. Psychologists at the National Research Council attempted to work out why some people were more "accident-prone" than others, and how driving tests and lessons could compensate for human failings. Some engineers — including former pilot Hugh DeHaven at Cornell University in Ithaca, New York — admitted that accidents were inevitable. They created the science of crashworthiness in the 1930s and 1940s. Here, the details become grisly. Experiments to work out how car design could protect human bodies began with dropping eggs from different heights and led to tests on live dogs and human cadavers (by the early 2000s, clad in "Smurf-blue leotards" to keep their extremities together after impact).

In the mid-twentieth century, the debate about road safety became more technocratic and lost some of its passion. It was reawakened in the 1960s by campaigners such as consumer advocate Ralph Nader, whose reputation was enhanced by the industry's

attempts to discredit him. Nader blended scientific evidence with political nous to argue that carmakers' disregard for safety was unjust as well as unwise. And under presidents Lyndon B. Johnson and Richard Nixon, the government began to assert itself on both car safety and environmental pollution. Manufacturers would no longer be able to add safety features just as a luxury. They were forced to develop technologies such as crumple zones to absorb impacts, and catalytic converters to meet tough laws imposed by the Environmental Protection Agency in the 1970s. Vinsel points out that the 2015 scandal in which Volkswagen admitted it had cheated on emissions tests has plenty of precedents from this period.

Vinsel's argument is that regulation involves the definition of problems. Scientific knowledge alone will not force action, and engineers need to be told what to focus on. In the history of car safety, problems have mostly been defined by carmakers, who have prioritized comfort over safety. Where they have focused on safety, the trade-offs have been problematic. The now-ubiquitous sports utility vehicle (SUV) — safer for drivers, but more likely to kill pedestrians — is a product of this view, and also a cautionary tale of unintended consequences. It was designed to be classed as a truck, and therefore exempt from emissions controls.

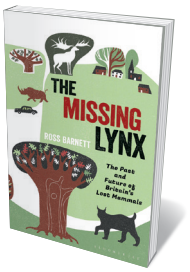
"Road safety started to be acknowledged as a problem but carmakers found it easy to offload responsibility."

Vinsel wants to be optimistic. He sees his story as a case of government regulation steering technologies in a positive direction. His book, however, is entirely US-centric, even though the market for cars is global and technological standards have been exported and imported. His argument could have been more powerful with some international comparisons.

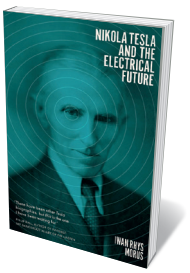
The US record on road safety remains woeful; the death rates per kilometre in Sweden and Britain are less than half those in the United States. Self-driving cars look like a poor technological fix for this problem. At a time when tech companies including Facebook, Uber and Google are given a free rein by US regulators and the specious promises of self-driving cars are used to justify further deregulation, a defence of government's role in technological development is much needed. ■

Jack Stilgoe *is an associate professor of science and technology studies at University College London. His forthcoming book* Who's Driving? *will be published by Palgrave Macmillan.*
e-mail: j.stilgoe@ucl.ac.uk

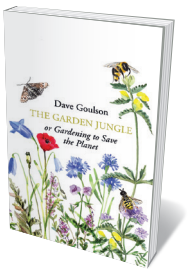
Books in brief



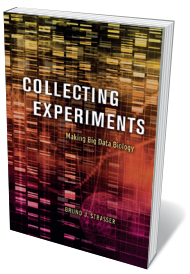
The Missing Lynx
Ross Barnett BLOOMSBURY *(2019)*
The story of life on Earth is a saga of extinction, declares palaeontologist Ross Barnett in this fresh and assured natural history of departed megafauna. Arguing that human 'overkill' was (with climate change) a major driver long before our population exploded in the Holocene epoch, Barnett uses Britain as a microcosm of the planetary record. Here are long-gone species such as the cave hyena (*Crocuta crocuta spelaean*), the fearsome scimitar-toothed cat (*Homotherium latidens*) and the northern lynx (*Lynx lynx lynx*); thrilling tales of discovery; and the vagaries of reintroduction. An often moving tribute to lost marvels.



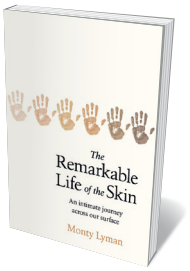
Nikola Tesla and the Electrical Future
Iwan Rhys Morus ICON *(2019)*
The Serbian inventor and electrical engineer Nikola Tesla seems to many uncannily prescient. Yet the scientist — by turns reclusive and flamboyant — was very much a product of the late nineteenth century. Historian Iwan Rhys Morus examines the man through that lens: a time of rampant entrepreneurialism, bravura innovations, grandiose visions of techno-utopia and futuristic science fiction. His crisply succinct, beautifully synthesized study brings to life Tesla, his achievements and failures (such as interplanetary communication), and the hopeful thrum of an era before world wars.



The Garden Jungle
Dave Goulson JONATHAN CAPE *(2019)*
Woodlice, earthworms, earwigs: a seething Serengeti lurks in many a back garden. Apiologist Dave Goulson's wonderful book encourages such richness by delivering solid science on garden wilding. Calling out today's cocktail of industrial pesticides as extreme in residential settings, he shows how robust plants and natural predators such as lacewings do the job sustainably. He extols the delights of eating roadkill, shows how to craft hoverfly habitats and advocates growing heritage crop varieties. Above all, Goulson demonstrates that the domestic nature reserve is the first step towards saving the planet.



Collecting Experiments
Bruno Strasser UNIVERSITY OF CHICAGO PRESS *(2019)*
We often think of big data as an explosive departure from the past. Science historian Bruno Strasser reveals it as part of a historic continuum. The sense of 'information overload' has existed since the Renaissance, and today's data tsunami emerged from two traditions in biology: natural-history collecting and the lab. Hybridized, they led to vast accumulations of knowledge. Strasser's case studies compel, from geneticists' 'museums' of maize (corn) varieties to a groundbreaking mine of digital data, the 1965 *Atlas of Protein Structure and Sequence*, coproduced by bioinformatics pioneer Margaret Dayhoff.



The Remarkable Life of the Skin
Monty Lyman BANTAM *(2019)*
Physician Monty Lyman peels back the science on human skin in this absorbing, fact-packed study. Dubbing it the "Swiss Army knife" of organs, Lyman examines skin as a barrier against trauma, a carrier of microbes, a matrix for nerve endings and a screen for the emotions. He reveals that structurally it is an "ideal foam", explores skin–gut communication, looks at medicinal tattooing and muses over ritual cleansing. Skin, he shows, is a thing of both surface and depth, a very visible yet personal part of ourselves that can become a target, too, of egregious attacks against difference. **Barbara Kiser**

Correspondence

Reject US anti–China sentiment

Open societies have benefited immeasurably from an influx of international scientists. The rapid exchange of ideas and expertise depends upon such movement, as do innovation and economic growth. It is therefore extremely worrying to read reports of Chinese scientists in the United States being treated unfairly (see *Nature* **571**, 157; 2019).

About one-quarter of US National Academy of Sciences members and one-quarter of US Nobel prizewinners were born abroad — and many more are children of immigrants. Three of the past five presidents of the Royal Society in London came from overseas, myself included: I'm an Indian-born US and British national.

Appropriate immigration controls, national security, local laws and ethical norms must all be taken seriously. But 'innocent until proven guilty' is an axiom of law throughout the civilized world. Guilt should never be assumed on the basis of national origin or religious belief, or on a perceived association. Actions such as the internment of US citizens of Japanese descent during the Second World War or the blacklisting of actual or alleged communists during the McCarthy era are now considered shameful episodes in US history.

We scientists must stand up for openness and fairness. Discriminating against someone because of their ethnicity, turning down a collaboration or refusing a visa for a conference on the grounds of nationality, or simply making someone feel unwelcome because they are an immigrant — these are all morally objectionable and practically counterproductive. Such behaviour must cease.

V. Ramakrishnan MRC Laboratory of Molecular Biology, Cambridge, UK.
ramak@mrc-lmb.cam.ac.uk

Eukaryote origin: 2D or not 2D?

The intense debate over the origin of eukaryotes is being fuelled by the lipids in the cell membrane of an engineered bacterium (see *Nature* **569**, 322–324; 2019). These could be evidence for a 'two-domain' (2D) model, in which eukaryotes diverged from a subgroup of the archaea. But in our view, the lipids offer better support for the 'three-domain' (3D) model, in which the two groups of organisms share a common ancestor.

Bacteria and eukaryotes have a similar set of lipids in their membranes. Archaeal membranes contain a different set. In both the 2D and 3D scenarios, there could have been an intermediate organism with mixed lipids. This might have arisen either during the transition from archaea to eukaryotes (2D) or at the start of the archaeal lineage (3D).

John van der Oost and his colleagues present an argument that could favour the 3D scenario — namely that the engineered bacterium, which also contains archaeal membrane lipids, is more resistant to heat shock than normal bacteria are (A. Caforio *et al. Proc. Natl Acad. Sci. USA* **115**, 3704–3709; 2018). This might explain why archaeal lipids were selected when archaeal ancestors started colonizing hot springs. The 2D model requires archaeal lipids to have been replaced in eukaryotes by weaker bacterial lipids — which seems to us implausible.

Patrick Forterre, **Morgan Gaia**, **Violette Da Cunha** *Pasteur Institute, Paris, France.*
patrick.forterre@pasteur.fr

Boost science input into SDGs

Government heads will meet in September to review progress on the United Nations' 17 Sustainable Development Goals (SDGs) and set a course for the next four years. The latest predictions are that no single country will meet all of the goals by the 2030 deadline (see go.nature.com/2nmfsxf) and that "countries need to step up efforts and act fast" (*Nature Sustain.* **1**, 377; 2018).

Universities, businesses and science academies are rallying their communities to make the SDGs a reality. An InterAcademy Partnership report released in May calls for the global science community, particularly national academies, to support the goals more effectively and with greater urgency (see go.nature.com/2xqxq73).

The report, entitled 'Improving Scientific Input to Global Policymaking with a focus on the UN Sustainable Development Goals', highlights mechanisms for feeding science into the UN and ways in which scientists can get involved. These include helping to strengthen targets and indicators for the SDGs, plugging data gaps and monitoring progress. Understanding interactions between goals, as well as their impact on wide-ranging policy interventions, is crucial. The report also advises on concrete actions that the scientific community should take to ensure that the best evidence is brought to bear at national, regional and global levels.

Scientists in all countries, from all disciplines and across all generations must play their part.

Teresa Stoepler, **Tracey Elliott** *The InterAcademy Partnership, Washington DC, USA.*
Eva Alisic *The University of Melbourne, Australia.*
telliott@iapartnership.org

EU farmers need independent advice

Pollution from nutrients and pesticides, degradation of soils and loss of habitats and biodiversity are a function of developments in agricultural science, as well as of farmers' management decisions (see G. Schmidt-Traub *et al. Nature* **569**, 181–183; 2019).

Farming practices have been transformed over the past half-century by specialization and simplification (for instance, the division of livestock and cash-crop farming) and by agrotechnologies such as biotechnology, robotics and remote sensing. However, there has been no parallel development of farming-systems science to integrate the short- and long-term economic, environmental and social effects of these innovative technologies at local scales. In Germany, for example, most technical advice for farmers is provided by representatives of the agricultural supply chain, who can neither evaluate their recommendations in a whole-farm context, nor assess the rebound effects on the biosphere.

The world's broken food system needs innovations in farming-systems science — for example, in the ecology of crop and livestock farming, in the biogeochemistry of land use and in the ethics of livestock husbandry and rural sociology. The benefits of new practices need to be realized without adverse environmental consequences.

Hans Schnyder* *Technical University of Munich, Germany.*
**On behalf of 4 correspondents (see go.nature.com/3jqmiv for full list).*
schnyder@wzw.tum.de

CONTRIBUTIONS

Correspondence may be submitted to **correspondence@nature.com** after consulting the author guidelines and section policies at **go.nature.com/cmchno**.

CLIMATE SCIENCE

Oddities in ocean record resolved

An analysis of the record of sea surface temperature reveals that some climate variations that are thought to have occurred in the North Atlantic and the North Pacific oceans are an artefact of changes in measurement approaches. SEE LETTER P.393

ZEKE HAUSFATHER

Reducing uncertainties in the historical record of Earth's surface temperature can improve scientists' ability to understand and explain changes in the climate over the past 150 years. This is particularly important for the early part of the twentieth century, because the cause of observed warming at that time remains fiercely debated[1]. On page 393, Chan *et al.*[2] demonstrate an innovative approach to account for differences in how sea surface temperature was measured in the early twentieth century. Their results suggest modestly less warming in the North Atlantic Ocean and substantially greater warming in the North Pacific Ocean during the period from 1908 to 1941, relative to previous estimates. Such findings indicate that intrinsic climate variability has a smaller impact on regional warming rates than was thought.

Improving historical temperature estimates has long been a key focus for climate researchers. Until the past few decades, most temperature measurements on both land and ocean were not aimed at detecting long-term climate changes. Rather, they were mainly intended to document average climate conditions or were for shorter-term meteorological purposes[3]. Adjustments to measurement methods that introduced biases of a few tenths of a degree Celsius were common. Although these biases were of little concern at the time, they become substantially more relevant when trying to detect long-term changes in global temperature of about 1 °C over the past 150 years.

The record of global surface temperature is produced by combining measurements of sea surface temperature (SST) with measurements of air temperature over land and ice. The largest remaining uncertainties in the global temperature record are associated with the SST estimates. Specifically, changes in observational instrumentation and techniques over time, coupled with patchy metadata (information about data) and sparse sampling in some regions complicate the interpretation of the historical record[4].

"The method offers an innovative solution to the lack of good ship metadata during the early twentieth century."

Initially, SST estimates were made using wooden buckets that were thrown over the sides of ships, filled with water and hauled up. The temperature of the water in the buckets was then measured using a thermometer. While the buckets were being hoisted up, evaporative cooling and exposure to ambient conditions would often reduce the temperature of the water by a few tenths of a degree Celsius.

This bias was exacerbated by a transition to poorly insulated canvas buckets in the late nineteenth century, and these buckets continued to be the main means of SST measurement until the period of the Second World War. Accounting for the cold bias in bucket measurements is the single largest adjustment to the ocean (and global) temperature record. Without the adjustment, the estimated rate of ocean warming from 1850 to the present would be about 30% higher[5].

A bucket measurement can be affected by a wide range of factors. These include the height of the ship, the composition and size of the bucket, how long it remains in the sea, whether the water is stirred before measurement and how long the thermometer is left in the water. Little of this information was recorded in a form that has survived to the present day. As a result, researchers have often had to inaccurately treat many bucket measurements as having the same magnitude of bias.

Chan and colleagues found a clever way to tackle this problem. They looked at the difference between SST measurements that were made within 300 kilometres and 2 days of one another, producing a data set of 6 million measurement pairs between 1908 and 1941. Ships were grouped by national origin, on the assumption that ships from the same country would tend to have similar measurement practices at any given time. The authors found sizeable offsets in SST estimates between ship

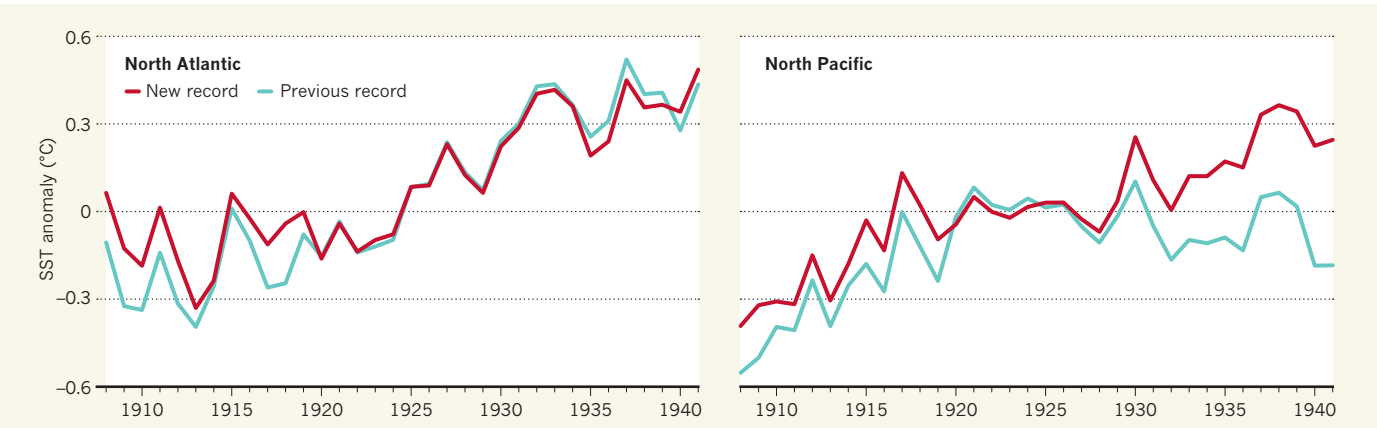


Figure 1 | Adjustments to sea surface temperature (SST) data. Chan *et al.*[2] propose corrections to the SST record of the North Atlantic and the North Pacific oceans from 1908 to 1941. The new record suggests slightly less warming in the North Atlantic and much greater warming in the North Pacific, compared with the previous record. The SST data are expressed as a departure (anomaly) from the average value during the period 1920–29. (Adapted from Fig. 4 of the paper[2].)

groups, ranging from −0.3 °C to +0.6 °C.

Digging further into these differences, Chan *et al.* realized that measurements from Japanese ships in the North Pacific suddenly became about 0.35 °C cooler after 1930 when compared with measurements from other countries. This change was caused by the Japanese switching from recording temperatures in whole-degrees Fahrenheit to taking readings in degrees Celsius and then dropping any numbers after the decimal point. The authors identified a similarly large change in the North Atlantic that is associated with German readings, but the cause of this change is less clear.

Chan and colleagues' results suggest that scientists have been overestimating warming in the North Atlantic and substantially underestimating warming in the North Pacific during the early twentieth century because of not fully accounting for biases in bucket measurements (Fig. 1). These findings bring the difference in estimated warming between the two regions in line with projections from climate models. However, there are still large differences between models and observations in the overall rate of global ocean warming during this period.

The authors' approach of comparing groups of proximate-ship measurements is conceptually similar to that used in identifying problems in the land temperature record, whereby each weather station is compared with its neighbours to find and remove localized biases[6]. The method offers an innovative solution to the lack of good ship metadata during the early twentieth century and provides a major advance in our understanding of historical ocean measurements.

This study, and recent major updates to the SST record at the UK Met Office's Hadley Centre[7], provide a useful reminder that large systematic biases might remain in our observational temperature records. Improved quantification of these biases is still a key technical challenge for researchers, and will help to address questions about the performance of climate-model simulations of the past and the role of intrinsic climate variability in historical temperature change. ■

Zeke Hausfather *is in the Energy and Resources Group, University of California Berkeley, Berkeley, California 94720, USA.*
e-mail: hausfath@berkeley.edu

1. Haustein, K. & Otto, F. E. L. *J. Clim.* https://doi.org/10.1175/JCLI-D-18-0555.1 (2019).
2. Chan, D., Kent, E. C., Berry, D. I. & Huybers, P. *Nature* **571**, 393–397 (2019).
3. Kent, E. C. *Bull. Am. Meteorol. Soc.* **98**, 1601–1616 (2017).
4. Lenssen, N. J. L. *et al. J. Geophys. Res. Atmos.* https://doi.org/10.1029/2018JD029522 (2019).
5. Kennedy, J. J., Rayner, N. A., Smith, R. O., Parker, D. E. & Saunby, M. *J. Geophys. Res. Atmos.* **116**, D14103 (2011).
6. Menne, M. J. & Williams, C. N. Jr. *J. Clim.* **22**, 1700–1717 (2009).
7. Kennedy, J. J., Rayner, N. A., Atkinson, C. P. & Killick, R. E. *J. Geophys. Res. Atmos.* https://doi.org/10.1029/2018JD029867 (2019).

GENETICS

How mutations express themselves

A method for detecting mutations and measuring gene-expression levels in the same cell has enabled an investigation into the effects of mutations in a specific gene on the emergence of a form of blood cancer. SEE ARTICLE P.355

SIDDHARTH RAJU & CHUN JIMMIE YE

The cells that circulate in the bloodstream perform various functions and, in adults, are derived from progenitor cells in the bone marrow. Mutations in the DNA sequences of progenitor cells can lead to changes in blood-cell development, sometimes resulting in cancer. Owing to technical constraints, elucidating the effects of progenitor mutations on blood-cell development has been challenging. On page 355, Nam *et al.*[1] report a method for detecting mutations and measuring gene expression in individual blood progenitor cells, and use it to analyse a mixture of progenitors with or without mutations in a cancer-linked gene. They show that progenitors that have the same mutation can give rise to cells with different gene-expression profiles.

Haematopoiesis — the process through which mature blood cells are formed from progenitors — is tightly regulated. The 'decision' that progenitor cells make as to which cell type to become is generally determined by the signals that they receive from their immediate surroundings. However, mutations that sometimes arise in these progenitor cells can result in the signals being blocked, overamplified or simply ignored, resulting in the enrichment or depletion of specific cell types and, in some cases, production of cancerous clones. Understanding how mutations in progenitor cells lead to changes in the production of different cell types is a key question.

Investigating how mutations in a progenitor cell affect its gene expression, and thus its identity and function, has been highly challenging, largely because mutant cells can be rare and often do not express molecular markers that can be used to separate them physically from non-mutant cells. Strategies to simultaneously detect genetic differences and measure gene expression in single cells have been used to assign cells from a mixture of immune blood cells to their human donor of origin[2], and to study changes in populations of host and

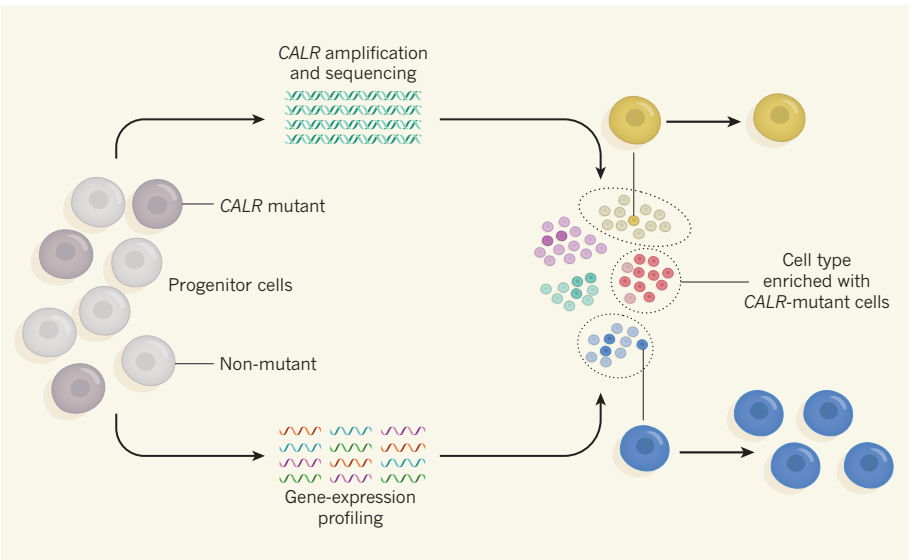


Figure 1 | An analysis of mutation status and gene expression in single cells. Nam *et al.*[1] sampled progenitor cells that give rise to blood cells from individuals who have a type of blood cancer that is caused by progenitor cells with mutations in the *CALR* gene. To distinguish mutant from non-mutant cells, the authors amplified and sequenced the *CALR* gene of individual cells. The authors also measured the levels of gene expression in each cell. They identified different cell types on the basis of a statistical analysis of the cells' gene-expression profiles (dotted circles represent statistical, rather than physical, cell groupings), and examined which of the cells in these different types had *CALR* mutations. Certain cell types were enriched in *CALR*-mutant cells, and *CALR* mutations had different effects (for example, on proliferation) in cells of different types.

https://doi.org/10.1038/d41586-019-02028-2

GENETICS

How mutations express themselves

A method for detecting mutations and measuring gene-expression levels in the same cell has enabled an investigation into the effects of mutations in a specific gene on the emergence of a form of blood cancer.

SIDDHARTH RAJU & CHUN JIMMIE YE

The cells that circulate in the bloodstream perform various functions and, in adults, are derived from progenitor cells in the bone marrow. Mutations in the DNA sequences of progenitor cells can lead to changes in blood-cell development, sometimes resulting in cancer. Owing to technical constraints, elucidating the effects of progenitor mutations on blood-cell development has been challenging. Writing in *Nature*, Nam *et al.*[1] report a method for detecting mutations and measuring gene expression in individual blood progenitor cells, and use it to analyse a mixture of progenitors with or without mutations in a cancer-linked gene. They show that progenitors that have the same mutation can give rise to cells with different gene-expression profiles.

Haematopoiesis — the process through which mature blood cells are formed from progenitors — is tightly regulated. The 'decision' that progenitor cells make as to which cell type to become is generally determined by the signals that they receive from their immediate surroundings. However, mutations that sometimes arise in these progenitor cells can result in the signals being blocked, overamplified or simply ignored, resulting in the enrichment or depletion of specific cell types and, in some cases, production of cancerous clones. Understanding how mutations in progenitor cells lead to changes in the production of different cell types is a key question.

Investigating how mutations in a progenitor cell affect its gene expression, and thus its identity and function, has been highly challenging, largely because mutant cells can be rare and often do not express molecular markers that can be used to separate them physically from non-mutant cells. Strategies to simultaneously detect genetic differences and measure gene expression in single cells have been used to assign cells from a mixture of immune blood cells to their human donor of origin[2], and to study changes in populations of host and donor cells in individuals with a type of blood cancer who received stem-cell transplants[3]. However, combined approaches have not been extensively used to examine the effects of mutations in cancer-associated genes on blood-cell development.

Nam *et al.* designed a method called 'genotyping of transcriptomes' (GoT) by combining an existing platform for profiling gene expression[3] with a technique for amplifying a specific genetic sequence to detect mutations in it (Fig. 1). They used this method to analyse thousands of progenitor cells sampled from the bone marrow of five individuals with a form of blood cancer that is caused by mutations in the *CALR* gene, and that is characterized by overproduction of platelet cells. GoT enabled the authors to ascertain which of the sampled cells carried a *CALR* mutation and which did not.

The authors used a statistical analysis to 'group' the sampled progenitor cells into different types on the basis of their gene-expression profiles (Fig. 1). All of the identified types contained both cells with and without the *CALR* mutation. However, *CALR*-mutant cells were more likely to follow certain differentiation pathways and therefore to become certain types of blood cell. Furthermore, Nam and colleagues found that the effects of the mutation, when present in the progenitor cells, were noticeable only at later stages of cellular differentiation; the progeny of *CALR*-mutant cells were more abundant than the progeny of their non-mutant counterparts and had a distinct gene-expression profile. Such observations would not have been possible using standard techniques, which demonstrates the value of this method.

Although GoT has its limitations, they can probably be addressed by adapting it to new single-cell workflows. First, GoT currently requires that the identity of the mutated gene, or a small set of potentially mutated genes, is known in advance. As an example, the authors used a multiplexed version of their analysis that can simultaneously target multiple prespecified parts of the genetic sequence to probe three genes. If no specific mutations, genes or regions of the genome have been prespecified for analysis (for example, on the basis of an association with disease progression), multiplexed analyses can, in theory, be used to cover larger panels of genes; however, this might not be cost-effective.

Second, GoT is less effective at detecting mutations that occur near the middle of a gene than those that occur near the ends. One

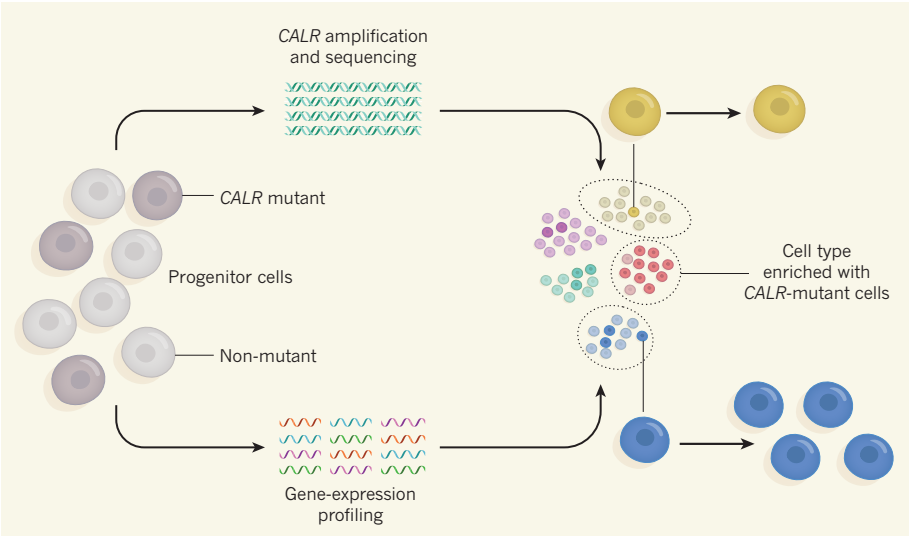


Figure 1 | An analysis of mutation status and gene expression in single cells. Nam *et al.*[1] sampled progenitor cells that give rise to blood cells from individuals who have a type of blood cancer that is caused by progenitor cells with mutations in the *CALR* gene. To distinguish mutant from non-mutant cells, the authors amplified and sequenced the *CALR* gene of individual cells. The authors also measured the levels of gene expression in each cell. They identified different cell types on the basis of a statistical analysis of the cells' gene-expression profiles (dotted circles represent statistical, rather than physical, cell groupings), and examined which of the cells in these different types had *CALR* mutations. Certain cell types were enriched in *CALR*-mutant cells, and *CALR* mutations had different effects (for example, on proliferation) in cells of different types.

solution to this problem would be to use a lower-throughput platform that allows the analysis of full-length RNA transcripts in single cells[4,5]; in theory, this approach could detect mutations anywhere in the RNA-encoding parts of genes. Nam *et al.* present an alternative approach by showing that a technique called nanopore sequencing, in which full-length transcripts are sequenced by passing them through a tiny pore, is compatible with their high-throughput platform.

Third, GoT cannot detect mutations in genetic sequences that are not transcribed but that may affect gene expression. Investigation of such sequences might be possible by combining GoT with a technique that measures how accessible certain DNA sequences in a cell are to enzymes[6].

A recent paper[7] used a different high-throughput approach to implement a similar targeted-amplification strategy to study a blood cancer that is thought to be partly caused by disruption of haematopoiesis by progenitor-cell mutations. The authors of that paper also identified a set of genes that were co-expressed only in malignant progenitors (that is, progenitor cells with a cancer-associated mutation), and described a machine-learning approach that used gene-expression data to distinguish malignant cells from non-malignant ones, even without using prespecified gene-sequence information. It would be interesting to see whether the same machine-learning approach could use Nam and colleagues' gene-expression data to distinguish the malignant cells from non-malignant cells. Obtaining gene-sequence information from single cells remains more challenging than assessing gene expression; therefore, a method for predicting malignancy solely on the basis of single-cell gene expression would have vast clinical implications.

In theory, GoT and similar approaches could be used to study any cancer. They have the potential to precisely determine the effects of mutations in known genes on downstream cell-development states and to establish whether certain mutations are sufficient to induce cancer. These insights, in turn, could shed light on the mechanisms that underlie the evolution of clonal lineages of cells in cancer. ■

Siddharth Raju *and* **Chun Jimmie Ye** *are in the Department of Medicine, Institute for Human Genetics, and at the Bakar Institute for Computational Health Sciences, University of California, San Francisco, San Francisco, California 94143, USA.*
e-mail: jimmie.ye@ucsf.edu

1. Nam, A. S. *et al. Nature* https://doi.org/10.1038/s41586-019-1367-0 (2019).
2. Kang, H. M. *et al. Nature Biotechnol.* **36**, 89–94 (2018).
3. Zheng, G. X. Y. *et al. Nature Commun.* **8**, 14049 (2017).
4. Gupta, I. *et al. Nature Biotechnol.* **36**, 1197–1202 (2018).
5. Macaulay, I. C. *et al. Nature Methods* **12**, 519–522 (2015).
6. Buenrostro, J. D. *et al. Nature* **523**, 486–490 (2015).
7. van Galen, P. *et al. Cell* **176**, 1265-1281 (2019).

50 Years Ago

This year is the bicentenary of the granting of patents for two inventions which played a crucial part in making Britain the most important nineteenth century industrial power. In 1769, James Watt patented his separate condenser, which proved to be the greatest single improvement ever made in steam engines, and Richard Arkwright patented his spinning machine, which, strictly speaking, was … a successful exploitation of a much earlier machine which never quite worked. To mark the occasion, the Science Museum in London has arranged a characteristically subdued exhibition of the two original patents … a little biographical material … and eight or nine cases containing recent and contemporary models and drawings of Watt's work and Arkwright's original spinning machines.

From *Nature* 19 July 1969

100 Years Ago

With the view of honouring some of those who helped to win the war … the North-East Coast Institution of Engineers and Shipbuilders held a Victory meeting … Lady Parsons read a paper on women's work in engineering and shipbuilding during the war. … There is no doubt that many women developed great mechanical skill and a real love of their work. The engineering industry is again barred to women by an agreement made between the Treasury and the trade unions … The meeting agreed with Lady Parson's condemnation of the Labour party, which, while demanding full political equality for women and their right to sit in the House of Lords and to practise at the Bar and as solicitors, will not grant to women equality of industrial opportunity.

From *Nature* 17 July 1919

donor cells in individuals with a type of blood cancer who received stem-cell transplants[3]. However, combined approaches have not been extensively used to examine the effects of mutations in cancer-associated genes on blood-cell development.

Nam *et al.* designed a method called 'genotyping of transcriptomes' (GoT) by combining an existing platform for profiling gene expression[3] with a technique for amplifying a specific genetic sequence to detect mutations in it (Fig. 1). They used this method to analyse thousands of progenitor cells sampled from the bone marrow of five individuals with a form of blood cancer that is caused by mutations in the *CALR* gene, and that is characterized by overproduction of platelet cells. GoT enabled the authors to ascertain which of the sampled cells carried a *CALR* mutation and which did not.

The authors used a statistical analysis to 'group' the sampled progenitor cells into different types on the basis of their gene-expression profiles (Fig. 1). All of the identified types contained both cells with and without the *CALR* mutation. However, *CALR*-mutant cells were more likely to follow certain differentiation pathways and therefore to become certain types of blood cell. Furthermore, Nam and colleagues found that the effects of the mutation, when present in the progenitor cells, were noticeable only at later stages of cellular differentiation; the progeny of *CALR*-mutant cells were more abundant than the progeny of their non-mutant counterparts and had a distinct gene-expression profile. Such observations would not have been possible using standard techniques, which demonstrates the value of this method.

Although GoT has its limitations, they can probably be addressed by adapting it to new single-cell workflows. First, GoT currently requires that the identity of the mutated gene, or a small set of potentially mutated genes, is known in advance. As an example, the authors used a multiplexed version of their analysis that can simultaneously target multiple prespecified parts of the genetic sequence to probe three genes. If no specific mutations, genes or regions of the genome have been prespecified for analysis (for example, on the basis of an association with disease progression), multiplexed analyses can, in theory, be used to cover larger panels of genes; however, this might not be cost-effective.

Second, GoT is less effective at detecting mutations that occur near the middle of a gene than those that occur near the ends. One solution to this problem would be to use a lower-throughput platform that allows the analysis of full-length RNA transcripts in single cells[4,5]; in theory, this approach could detect mutations anywhere in the RNA-encoding parts of genes. Nam *et al.* present an alternative approach by showing that a technique called nanopore sequencing, in which full-length transcripts are sequenced by passing them through a tiny pore, is compatible with their high-throughput platform.

Third, GoT cannot detect mutations in genetic sequences that are not transcribed but that may affect gene expression. Investigation of such sequences might be possible by combining GoT with a technique that measures how accessible certain DNA sequences in a cell are to enzymes[6].

A recent paper[7] used a different high-throughput approach to implement a similar targeted-amplification strategy to study a blood cancer that is thought to be partly caused by disruption of haematopoiesis by progenitor-cell mutations. The authors of that paper also identified a set of genes that were co-expressed only in malignant progenitors (that is, progenitor cells with a cancer-associated mutation), and described a machine-learning approach that used gene-expression data to distinguish malignant cells from non-malignant ones, even without using prespecified gene-sequence information. It would be interesting to see whether the same machine-learning approach could use Nam and colleagues' gene-expression data to distinguish the malignant cells from non-malignant cells. Obtaining gene-sequence information from single cells remains more challenging than assessing gene expression; therefore, a method for predicting malignancy solely on the basis of single-cell gene expression would have vast clinical implications.

"Understanding how mutations in progenitor cells lead to changes in the production of different cell types is a key question."

In theory, GoT and similar approaches could be used to study any cancer. They have the potential to precisely determine the effects of mutations in known genes on downstream cell-development states and to establish whether certain mutations are sufficient to induce cancer. These insights, in turn, could shed light on the mechanisms that underlie the evolution of clonal lineages of cells in cancer. ■

Siddharth Raju *and* **Chun Jimmie Ye** *are in the Department of Medicine, Institute for Human Genetics, and at the Bakar Institute for Computational Health Sciences, University of California, San Francisco, San Francisco, California 94143, USA.*
e-mail: jimmie.ye@ucsf.edu

1. Nam, A. S. *et al. Nature* **571**, 355–360 (2019).
2. Kang, H. M. *et al. Nature Biotechnol.* **36**, 89–94 (2018).
3. Zheng, G. X. Y. *et al. Nature Commun.* **8**, 14049 (2017).
4. Gupta, I. *et al. Nature Biotechnol.* **36**, 1197–1202 (2018).
5. Macaulay, I. C. *et al. Nature Methods* **12**, 519–522 (2015).
6. Buenrostro, J. D. *et al. Nature* **523**, 486–490 (2015).
7. van Galen, P. *et al. Cell* **176**, 1265-1281 (2019).

This article was published online on 3 July 2019.

https://doi.org/10.1038/d41586-019-02016-6

IMAGING TECHNIQUES

Light scatters electrons to make holograms

The quantum interference of electrons that have been scattered by light has been used to produce holograms of the underlying electromagnetic fields — and might open up methods for studying materials at high temporal and spatial resolution.

CLAUS ROPERS

If you ask people what a hologram is, they'll probably describe the 3D light projections of science-fiction films — such as the vision of Princess Leia floating in free space in the 1977 film *Star Wars*. Such projections are becoming a reality[1], but the original goal of holography in science is arguably more mundane: to record a property of wave fields known as the phase, which defines the pattern of peaks and troughs of a travelling wave at a given moment in time. For many physicists, this concept is just as exciting as a sci-fi hologram. Writing in *Science Advances*, Madan *et al.*[2] report new types of hologram produced by the scattering of electrons by light fields. Not only do these findings broaden the scope of electron holography, but they also allow both the amplitude and the phase of electromagnetic (light) waves to be determined.

Holography is a widely used measurement technique in electron microscopy that makes use of the wave character of electrons[3]. In this technique, two parts of an electron beam are overlapped to create a stripy interference pattern (the hologram). The difference in the phases of the two beams can be extracted from this pattern. Because electrical and magnetic fields can affect the phases of electron beams that pass through them, holography in electron microscopy can be used to quantitatively map such fields with extremely high spatial resolution, down to the nanometre scale.

However, Madan *et al.* wanted to measure the phase of oscillating light waves. To this end, they developed a holographic method that depends on a different principle from that used in conventional electron microscopy: the modulation of quantum interference between electrons by oscillating light fields. Let's consider the physical mechanism by which electrons interact with light in the authors' experiments.

If a stream of fast electrons traverses an oscillating electromagnetic field, some electrons accelerate whereas others decelerate, depending on when they enter and exit the field. Measurements of the velocity distribution of electrons that have passed through such a field have revealed that electrons pick up or lose energy in quantized amounts — specifically, in multiples of the energy of the photons in the light field[4,5]. The size of the effect increases with the light intensity, a relationship that is used as the basis of a technique called photon-induced near-field electron microscopy (PINEM) to map light intensities around nanoparticles and other small structures[6].

To measure the phase of light waves in PINEM experiments, some form of interference must be produced. Madan *et al.* created such interference by conducting PINEM experiments on samples that had different geometries, and in which the electrons interacted with more than one light wave. Some of these implementations involved electrons sequentially flying through two spatially separated light fields. As has been shown previously[7], the relative phase of these two fields determines the strength of the combined electron–light interactions: in-phase fields can enhance the interaction in a kind of constructive interference, whereas the two fields can cancel each other out if they have opposite phases.

In perhaps the most striking experiment of the paper, Madan *et al.* illuminated an aperture in a metallic film to produce waves called surface plasmon polaritons (SPPs), which are, essentially, light fields bound to the metal surface (Fig. 1). The electron beam passed through and interacted with these SPPs and with fields on the opposite side of the film. This created a spiral-shaped interference pattern that encoded the relative phases of the light fields at each position on the film, and therefore contained holographic information. When the authors tilted the film in the electron beam, the spiral became distorted in a way that reflected the different propagation directions of the SPPs — in much the same way as the pitch of an ambulance siren sounds different depending on whether the vehicle is driving towards or away from you.

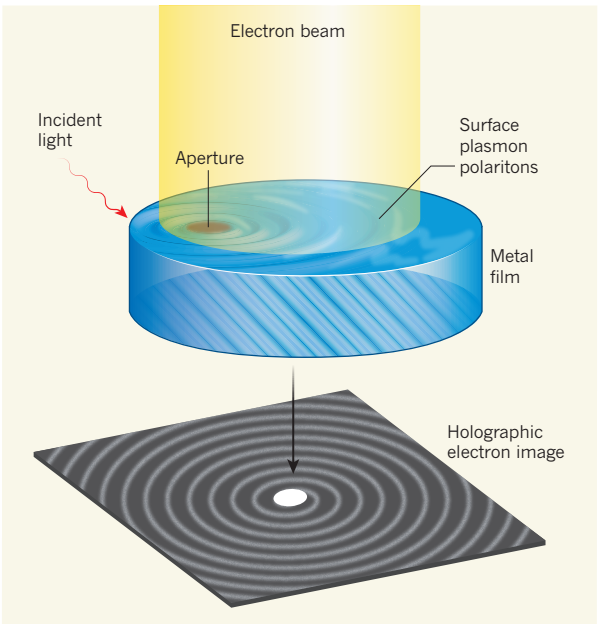


Figure 1 | Imaging of electron–light interference. Madan *et al.*[2] report new types of hologram produced by the scattering of electrons by light fields. In one of the experiments, light irradiates a metal film that contains an aperture, to produce waves called surface plasmon polaritons — light fields bound to the metal surface. A different light-field pattern (illustrated by stripes) is produced on the other side of the film. When an electron beam passes through the film, it subsequently interacts with the fields on both sides, producing a spiral interference pattern. This pattern encodes the relative phases (the patterns of peaks and troughs) of the light fields at each position on the film, and therefore contains holographic information.

Similar interference patterns have previously been observed[8] in experiments in which light fields lead to the emission of electrons from surfaces, but there are key conceptual differences between those experiments and the current work. Specifically, some of the holograms in the present study arise from the quantum-mechanical interference of electron beams, rather than from the interference of crossed light waves. Notably, in Madan and colleagues' study, the electrons, effectively, mediate interference between light waves that do not overlap. In other words, optical phase information is imprinted on an electron in one place and then 'read out' by a second light field at a different location.

The ability to use electrons to transport optical phase information potentially opens up a variety of applications in electron microscopy and

beyond. For example, it should be possible to use such phase information to measure the optical response of single or coupled quantum light emitters embedded in solids, such as individual atoms, molecules or point defects in a crystal. Getting electrons to interact with more-complicated laser-pulse sequences than in the current experiment, and with multiple colours of light, might facilitate entirely new forms of electron spectroscopy. Combined with methods for the light-induced temporal structuring of electron beams[9–11], Madan and colleagues' holographic approaches could enable the behaviour of materials to be studied on shorter timescales than that of a single wave cycle of light (the attosecond scale), and with the spatial resolution of an electron microscope.

It remains to be seen whether more-ambitious applications of the new findings will materialize, in which electron beams are used as part of quantum communication systems, or even in quantum computation. Such technologies would probably require the controlled coupling or quantum correlation of multiple free electrons with each other, neither of which has been achieved so far. In the meantime, Madan and colleagues' work represents exciting progress in the manipulation of electrons by light. ■

Claus Ropers *is at the IV. Physical Institute, University of Göttingen, 37077 Göttingen, Germany.*
e-mail: cropers@gwdg.de

1. Smalley, D. E. *et al. Nature* **553**, 486–490 (2018).
2. Madan, I. *et al. Sci. Adv.* **5**, eaav8358 (2019).
3. Lichte, H. & Lehmann, M. *Rep. Prog. Phys.* **71**, 016102 (2008).
4. García de Abajo, F. J., Asenjo-Garcia, A. & Kociak, M. *Nano Lett.* **10**, 1859–1863 (2010).
5. Park, S. T., Lin, M. & Zewail, A. H. *New J. Phys.* **12**, 123028 (2010).
6. Barwick, B., Flannigan, D. J. & Zewail, A. H. *Nature* **462**, 902–906 (2009).
7. Echternkamp, K. E., Feist, A., Schäfer, S. & Ropers, C. *Nature Phys.* **12**, 1000–1004 (2016).
8. Spektor, G. *et al. Science* **355**, 1187–1191 (2017).
9. Priebe, K. E. *et al. Nature Photon.* **11**, 793–797 (2017).
10. Kozák, M., Schönenberger, N. & Hommelhoff, P. *Phys. Rev. Lett.* **120**, 103203 (2018).
11. Morimoto, Y. & Baum, P. *Nature Phys.* **14**, 252–256 (2018).

crystal. Getting electrons to interact with more-complicated laser-pulse sequences than in the current experiment, and with multiple colours of light, might facilitate entirely new forms of electron spectroscopy. Combined with methods for the light-induced temporal structuring of electron beams[9–11], Madan and colleagues' holographic approaches could enable the behaviour of materials to be studied on shorter timescales than that of a single wave cycle of light (the attosecond scale), and with the spatial resolution of an electron microscope.

It remains to be seen whether more-ambitious applications of the new findings will materialize, in which electron beams are used as part of quantum communication systems, or even in quantum computation. Such technologies would probably require the controlled coupling or quantum correlation of multiple free electrons with each other, neither of which has been achieved so far. In the meantime, Madan and colleagues' work represents exciting progress in the manipulation of electrons by light. ■

Claus Ropers *is at the IV. Physical Institute, University of Göttingen, 37077 Göttingen, Germany.*
e-mail: cropers@gwdg.de

1. Smalley, D. E. *et al. Nature* **553**, 486–490 (2018).
2. Madan, I. *et al. Sci. Adv.* **5**, eaav8358 (2019).
3. Lichte, H. & Lehmann, M. *Rep. Prog. Phys.* **71**, 016102 (2008).
4. García de Abajo, F. J., Asenjo-Garcia, A. & Kociak, M. *Nano Lett.* **10**, 1859–1863 (2010).
5. Park, S. T., Lin, M. & Zewail, A. H. *New J. Phys.* **12**, 123028 (2010).
6. Barwick, B., Flannigan, D. J. & Zewail, A. H. *Nature* **462**, 902–906 (2009).
7. Echternkamp, K. E., Feist, A., Schäfer, S. & Ropers, C. *Nature Phys.* **12**, 1000–1004 (2016).
8. Spektor, G. *et al. Science* **355**, 1187–1191 (2017).
9. Priebe, K. E. *et al. Nature Photon.* **11**, 793–797 (2017).
10. Kozák, M., Schönenberger, N. & Hommelhoff, P. *Phys. Rev. Lett.* **120**, 103203 (2018).
11. Morimoto, Y. & Baum, P. *Nature Phys.* **14**, 252–256 (2018).

This article was published online on 3 July 2019.

COMPUTATIONAL CHEMISTRY

Holistic models of reaction selectivity

Computational models that predict the selectivity of reactions are typically accurate for only a specific reaction type and a narrow range of reaction components. A more general model has now been reported. SEE ARTICLE P.343

PER-OLA NORRBY

Selectivity is a linchpin of chemical synthesis — if a synthetic reaction is not selective, it cannot give a good yield of the desired product, and will require tedious purification processes. Chemists have therefore long sought ways of predicting the selectivity of chemical reactions. Computational models can be constructed, but their development is laborious, and they are usually specific to a particular reaction type. On page 343, Reid and Sigman[1] now show that a selectivity model can be built in a semi-automated way and generalized over a range of reactions.

Chemical selectivity comes in many flavours, but it is especially difficult to achieve enantioselectivity, which depends on a property called chirality. Molecules are said to be chiral if they come as two mirror-image forms — enantiomers — that have many identical properties, but can differ in certain important aspects. A good analogy is with hands: a person's right and left hands have the same length, colour and mass, but only one fits into a right-handed glove.

Many biological targets for pharmaceuticals look like right-handed gloves to molecules — only one enantiomer of a molecule will fit into them. For this reason, pharmaceuticals should be synthesized as one enantiomer only; the other form might even be toxic. Asymmetrical catalysts are used to influence synthetic chemical reactions to form only one enantiomer of the product. Nature's asymmetrical catalysts are enzymes, which produce single enantiomers of biomolecules efficiently and with exquisite selectivity. Enzymes can also be used as catalysts for synthetic chemistry, but they generally have a limited range of substrates and can produce only one of the two possible enantiomers of a product.

Modern synthetic catalysts challenge the efficiency of enzymes, and can often be made as mirror-image forms that each produce a different enantiomer of a desired molecule. To support the development of new catalysts, chemists use models to understand and predict the enantioselectivity of catalytic reactions[2,3]. These range in complexity from simple models of the catalyst drawn on paper, onto which a molecular model of the substrate is superimposed to estimate the best fit, to quantum-mechanical calculations that describe an entire reaction path.

"It is highly encouraging to see that holistic reaction models can be produced by using a wide training set."

A direct predecessor of Reid and Sigman's modelling work is a computational approach called quantitative structure–selectivity relationships (QSSR), in which a correlation is sought between the properties of reaction components and the observed selectivity. The relevant properties can be either determined experimentally or calculated, and can include such things as molecular-bond lengths, vibrational frequencies and atomic charges. Using a semi-automated statistical approach (multiple linear regression), these properties are used to construct a model that outputs one numeric value for each reaction system being studied[3]. A result of zero means that there is no selectivity — both enantiomers are produced in equal amounts. A high value indicates a very selective system, and the sign of the numerical output (positive or negative) indicates which enantiomer is mostly produced. Opposite enantiomers of a catalyst produce opposite enantiomers of the product, and this should also be reflected in QSSR models of synthetic catalysts; this requirement is not essential for models of enzymes, however, because only one enantiomeric form of any enzyme exists in nature.

QSSR models are normally limited to a narrow set of substrates and catalysts, because the assumptions built into the machine-learning procedures are invalidated by large deviations from the molecular structures used to train the model. Reid and Sigman have taken on the challenge of making a general QSSR model, starting from an earlier model reported by Reid and colleagues[4].

Inspired by enzyme models, Reid and Sigman ignored the sign conventions usually adhered to in models of synthetic catalysis — that is, they produced a model that predicts the magnitude of enantioselectivity for a group of catalytic reactions (Fig. 1), but only for one enantiomer of the catalyst. Switching the catalyst to its mirror image will therefore not switch the sign of the output in their model, and the model cannot predict which enantiomer is produced as the major isomer. However, the major enantiomer can be predicted from the preceding work[4]. Within this framework, the authors demonstrated that one of the components of the modelled reactions could be varied to an unprecedented degree, without affecting the high accuracy of the predictions.

How can one model achieve such a wide range of accurate predictions? Part of the explanation is probably that all the reactions share a similar mechanism: a planar substrate (an imine molecule; Fig. 1) is 'gripped' from

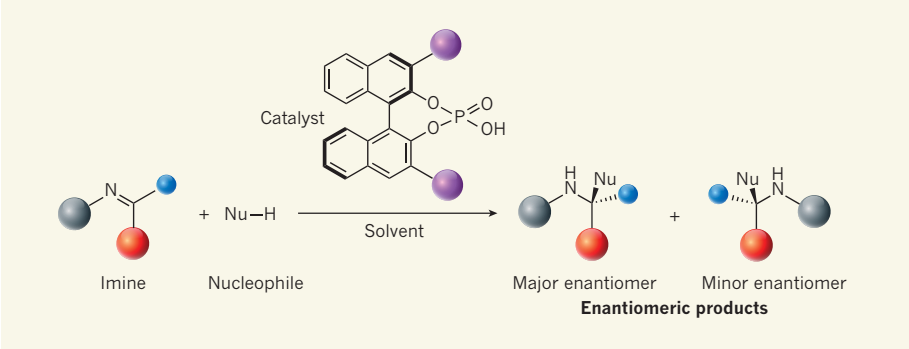


Figure 1 | Model reactions. Reid and Sigman[1] report a computational model that predicts the outcome of reactions when a wide range of nucleophilic molecules react with imines in the presence of a catalyst, accounting for factors such as molecular structure and solvent. More specifically, the model reports the magnitude of the enantioselectivity of the reactions — a measure of the ratio of the two mirror-image isomers (enantiomers) of the product formed in the reaction. Spheres represent a variety of chemical groups; bonds shown in bold or as solid wedges project above the plane of the page; broken wedges project below the plane of the page. Nu represents a range of groups or molecular structures.

one side by the chiral catalyst[4], so that any reaction has to occur on the other side. The third reaction component (a nucleophile), can therefore be varied substantially in the model. But the main reason is that the authors made a huge effort to produce a comprehensive training set of 367 individual reactions, each of which required multiple calculations to describe all the components, including the variability in shape (the conformations) of each component. It is highly encouraging to see that holistic reaction models can be produced by using such a wide training set.

Where next? A dream for reactivity modellers is to build an ultimate tool that accurately predicts the products of any reaction from the reaction components, thereby allowing computational screening of new reactions. Modellers have a long way to go to achieve this, but Reid and Sigman have shown that they can accurately predict outcomes for groups of related reactions, rather than having to model one type of reaction at a time. Other machine-learning methods are being tested on even bigger data sets[5].

The broadening of reaction scope demonstrated in the current work will encourage the search for more-general models, and might eventually enable models that predict the outcomes of reactions very different from those used for training. For now, making such predictions is still the domain of humans, but synthetic chemists will increasingly rely on theoretical tools to guide their work. I, for one, look forward to a future in which the tedious trial and error of synthetic chemistry is removed, and in which chemists can cut to the chase by carrying out only successful reactions. ■

Per-Ola Norrby *is in Data Science & Modelling, Pharmaceutical Sciences, R&D, AstraZeneca Gothenburg, 43183 Mölndal, Sweden.*
e-mail: per-ola.norrby@astrazeneca.com

1. Reid, J. P. & Sigman, M. S. *Nature* **571**, 343–348 (2019).
2. Brown, J. M. & Deeth, R. J. *Angew. Chem. Int. Ed.* **48**, 4476–4479 (2009).
3. Reid, J. P. & Sigman, M. S. *Nature Rev. Chem.* **2**, 290–305 (2018).
4. Reid, J. P., Simón, L. & Goodman, J. M. *Acc. Chem. Res.* **49**, 1029–1041 (2016).
5. Segler, M. H. S., Preuss, M. & Waller, M. P. *Nature* **555**, 604–610 (2018).

EVOLUTION

A deep dive into sea-squirt development

An analysis of gene expression in sea-squirt embryos at different stages of development deepens our understanding of how the body plans of vertebrates might have evolved from those of less complex animals. SEE ARTICLE P.349

NORIYUKI SATOH

Sea squirts such as *Ciona intestinalis* are the closest living invertebrate relatives of vertebrates. Their tadpole-like larvae feature some of the same organs and tissues as those found in developing vertebrates. On page 349, Cao *et al.*[1] use gene-expression data to examine the embryonic development of *C. intestinalis* larvae and to compare its development with that of other chordate animals, including vertebrates and cephalochordates, to reveal fresh insights into the evolution of vertebrates.

Single-cell analyses of gene expression have revolutionized various biological subdisciplines[2]. Such analyses at different stages of embryonic development have revealed how cells give rise to the various cell types that perform distinct functions and make up specific parts of the embryo[3,4]. As examples, studies of frog and zebrafish embryos have demonstrated that the three layers of cells that form these embryos — the ectoderm, endoderm and mesoderm — contain at least 50 cell types that have similar gene-expression profiles[3,4]. Studies into how different species develop often unveil clues to their evolutionary origins.

There are several advantages to studying embryonic development in sea squirts — which are also known as ascidians. As the closest relatives of vertebrates, they provide a reference for understanding the evolution of vertebrate body plans (Fig. 1). In *C. intestinalis*, embryogenesis — that is, the period of development that begins when cells are initially reorganized into a multilayered body of cells called a gastrula, and ends with larval hatching — takes just a day to complete. A *Ciona* larva comprises only about 2,500 cells, which make up distinctly differentiated organs and systems, including bilateral muscle, the central nervous system (CNS) and the notochord — a rod-like structure that gives rise to the backbone in vertebrates, and which is a defining characteristic of all chordate animals.

The cell lineages that comprise ascidian embryos have long been described[5]; the developmental fate of cells is restricted early in embryogenesis, at around the 110-cell stage. The *C. intestinalis* genome has been sequenced[6], and a network of genes and regulatory molecules that provides the blueprint for the body plan of all chordate animals has been characterized in *C. intestinalis*[7].

Cao *et al.* profiled the gene expression of more than 90,000 single cells from *C. intestinalis* at 10 developmental stages, from gastrulae to swimming larvae. The authors used these gene-expression data — carefully considering the expression of molecular markers of different cell types and lineages — to construct developmental trajectories of individual cell types. Whereas the larvae of *C. intestinalis* were previously thought to have approximately 20 cell types[8], Cao and colleagues' analysis identified 60 distinct cell types. A similarly comprehensive profiling of larval and embryonic cell types in vertebrates

https://doi.org/10.1038/d41586-019-01967-0

EVOLUTION

A deep dive into sea-squirt development

An analysis of gene expression in sea-squirt embryos at different stages of development deepens our understanding of how the body plans of vertebrates might have evolved from those of less complex animals.

NORIYUKI SATOH

Sea squirts such as *Ciona intestinalis* are the closest living invertebrate relatives of vertebrates. Their tadpole-like larvae feature some of the same organs and tissues as those found in developing vertebrates. Writing in *Nature*, Cao *et al.*[1] use gene-expression data to examine the embryonic development of *C. intestinalis* larvae and to compare its development with that of other chordate animals, including vertebrates and cephalochordates, to reveal fresh insights into the evolution of vertebrates.

Single-cell analyses of gene expression have revolutionized various biological subdisciplines[2]. Such analyses at different stages of embryonic development have revealed how cells give rise to the various cell types that perform distinct functions and make up specific parts of the embryo[3,4]. As examples, studies of frog and zebrafish embryos have demonstrated that the three layers of cells that form these embryos — the ectoderm, endoderm and mesoderm — contain at least 50 cell types that have similar gene-expression profiles[3,4]. Studies into how different species develop often unveil clues to their evolutionary origins.

There are several advantages to studying embryonic development in sea squirts — which are also known as ascidians. As the closest relatives of vertebrates, they provide a reference for understanding the evolution of vertebrate body plans (Fig. 1). In *C. intestinalis*, embryogenesis — that is, the period of development that begins when cells are initially reorganized into a multilayered body of cells called a gastrula, and ends with larval hatching — takes just a day to complete. A *Ciona* larva comprises only about 2,500 cells, which make up distinctly differentiated organs and systems, including bilateral muscle, the central nervous system (CNS) and the notochord — a rod-like structure that gives rise to the backbone in vertebrates, and which is a defining characteristic of all chordate animals.

The cell lineages that comprise ascidian embryos have long been described[5]; the developmental fate of cells is restricted early in embryogenesis, at around the 110-cell stage. The *C. intestinalis* genome has been sequenced[6], and a network of genes and regulatory molecules that provides the blueprint for the body plan of all chordate animals has been characterized in *C. intestinalis*[7].

Cao *et al.* profiled the gene expression of more than 90,000 single cells from *C. intestinalis* at 10 developmental stages, from gastrulae to swimming larvae. The authors used these gene-expression data — carefully considering the expression of molecular markers of different cell types and lineages — to construct developmental trajectories of individual cell types. Whereas the larvae of *C. intestinalis* were previously thought to have approximately 20 cell types[8], Cao and colleagues' analysis identified 60 distinct cell types. A similarly comprehensive profiling of larval and embryonic cell types in vertebrates and cephalochordates would not currently be feasible.

Vertebrates and their sister group, the urochordates — which include the ascidians — are thought to share a common ancestor with cephalochordate animals, such as amphioxus (Fig. 1a). Cao and colleagues' study provides at least two insights into the evolution of vertebrates from this common ancestor: one concerning the notochord, and the other concerning the CNS, which becomes especially complex in vertebrates.

Amphioxus larvae are fish-like, and their notochord consists of stiff, coin-shaped muscle cells. By contrast, the notochords of ascidian larvae and vertebrates lack muscle-like properties, and instead consist of cells containing fluid-filled vacuoles that provide stiffness for muscle-driven movements of the tail. How these distinct notochord types evolved has been unclear. Cao *et al.* provide gene-expression evidence that the *C. intestinalis* notochord exhibits properties of both types. Specifically, the anterior part of the notochord is typical of that of ascidians and vertebrates, whereas the posterior part consists of cells that have muscular properties, as in amphioxus larvae. However, how *C. intestinalis* produces the

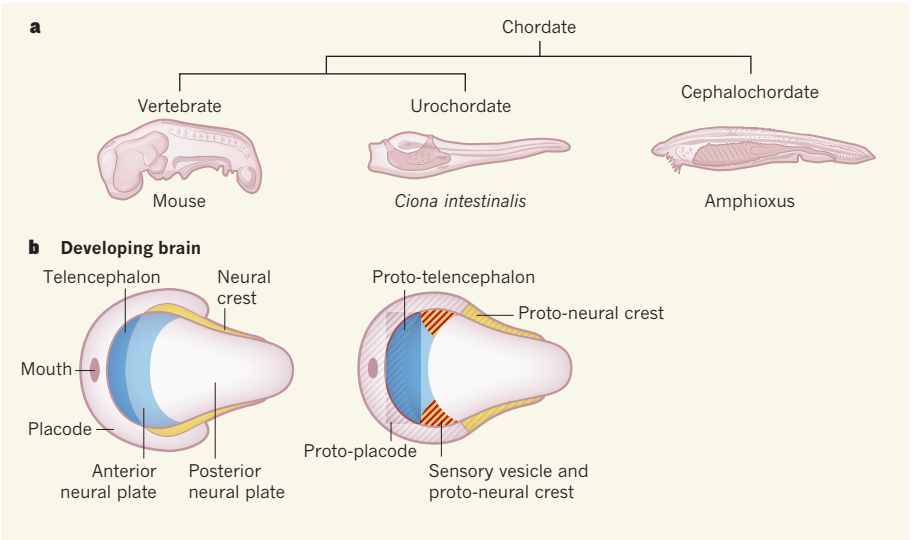


Figure 1 | Evolution of the chordate body plan. a, Vertebrates evolved from a common ancestor shared with urochordates, such as ascidians (including the sea squirt *Ciona intestinalis*), and the cephalochordate amphioxus. **b**, Whereas the amphioxus has no structures comparable to the vertebrate brain, cells in the sensory vesicle of the developing *C. intestinalis* nervous system express sets of regulatory genes that are also expressed in the neural crest and placode (structures of the developing vertebrate nervous system), suggesting that the simple brain of ascidian larvae contains prototypes of these regions. Cao *et al.*[1] measured gene expression in single cells of developing *C. intestinalis* embryos, and describe a cell-lineage map that reveals a prototype of the telencephalon (a part of the vertebrate brain that, in more complex vertebrates, is required for cognition) at the front of the larval ascidian brain. The neural plate gives rise to the brain and the spinal cord (in vertebrates) or the nerve cord (in urochordates). The position of the eventual mouth is also shown. In both panels, anterior is left, and posterior is right. (Embryo images adapted from ref. 15; CC BY 4.0.)

anterior and posterior parts independently and combines them into a single organ is not clear[9].

According to the 'new head' hypothesis[10], the evolution of vertebrates can be largely ascribed to the emergence of placodes and the neural crest, which are developmental populations of cells that give rise to most of the tissues in the head and jaw. Previous evolutionary–developmental studies have shown that *C. intestinalis* possesses rudimentary versions of these two key vertebrate innovations[11,12]. In contrast to the CNS of amphioxus larvae, which is not very well organized, the CNS of ascidian larvae resembles a prototype of the vertebrate brain (Fig. 1b). Cao and colleagues identified 41 neural cell types in *C. intestinalis* larvae, including peripheral sensory cells and interneuronal cells, and showed that each type mapped to a specific region of the CNS, including the sensory vesicle (the anterior part of the CNS in urochordates), the motor ganglion (a cluster of neurons that control movement) and the nerve cord (the bundle of neuronal fibres that runs along the length of the body of chordate animals).

Cao and co-workers' data also help to elucidate the evolutionary origins of the telencephalon of the vertebrate brain; in many higher vertebrates, the telencephalon is enlarged and is crucial for perception and cognition. The gene-expression profiles and developmental trajectories of cells in the anterior-most regions of the neural plate (a developmental structure that gives rise to the CNS) revealed that these regions, particularly the sensory cells of the palps (protrusions of ectoderm tissue at the front of the larva) and the pro-anterior sensory vesicle (located at the anterior of the developing nervous system) are the invertebrate counterparts of the vertebrate telencephalon. The vertebrate telencephalon is thus likely to have arisen through the incorporation of non-neural ectoderm into anterior regions of the developing nervous system.

The finding of a prototype of the vertebrate telencephalon in the ascidian larva raises the question of how the complicated structure of the vertebrate brain evolved. Similar genetic and developmental trajectories are probably shared by ascidians and vertebrates, but the much more complex architecture of vertebrate brains means that they can perform more-sophisticated functions. Further experiments using single-cell analysis on amphioxus larvae will be needed to help determine how the complex architecture of the vertebrate brain arose.

Over the past two decades, there have been great advances in our understanding of the molecular, cellular and developmental mechanisms involved in the origins of chordates[13] and the evolution of vertebrates[14]. As shown by Cao *et al.*[1], single-cell analyses of gene expression deepen our understanding of the evolutionary emergence of cell types that confer vertebrate-specific properties. This line of research also highlights the importance of genomic information and the wide-ranging scope of analyses of gene regulatory networks. Mechanisms of vertebrate evolution revealed by evolutionary–developmental studies will increasingly be based on detailed and precise data from gene regulatory networks in individual cells, tissues and organs, and data acquired using other new techniques, such as those that probe the architecture of DNA complexes in the nuclei of individual cells, and sophisticated computational tools. ■

Noriyuki Satoh *is in the Marine Genomics Unit, Okinawa Institute of Science and Technology Graduate University, Onna, Okinawa 904-0495, Japan.*
e-mail: norisky@oist.jp

1. Cao, C. *et al. Nature* https://doi.org/10.1038/s41586-019-1385-y (2019).
2. Stuart, T. & Satija, R. *Nature Rev. Genet.* **20**, 257–272 (2019).
3. Farrell, J. A. *et al. Science* **360**, eaar3131 (2018).
4. Briggs, J. A. *et al. Science* **360**, eaar5780 (2018).
5. Conklin, E. G. *J. Acad. Nat. Sci. Philadelphia* **13**, 1–119 (1905).
6. Dehal, P. *et al. Science* **298**, 2157–2167 (2002).
7. Imai, K. S., Levine, M., Satoh, N. & Satou, Y. *Science* **312**, 1183–1187 (2006).
8. Satoh, N. *Developmental Genomics of Ascidians* (Wiley-Blackwell, 2014).
9. Harder, M., Reeves, W., Byers, C., Santiago, M. & Veeman, M. *Dev. Biol.* **448**, 136–146 (2019).
10. Gans, C. & Northcutt, R. G. *Science* **220**, 268–273 (1983).
11. Stolfi, A., Ryan, K., Meinertzhagen, I. A. & Christiaen, L. *Nature* **527**, 371–374 (2015).
12. Horie, R. *et al. Nature* **560**, 228–232 (2018).
13. Satoh, N. *Chordate Origins and Evolution* (Academic, 2016).
14. Gee, H. *Across the Bridge: Understanding the Origin of the Vertebrates*. (Univ. Chicago Press, 2018).
15. Inoue, J. & Satoh, N. *Mol. Biol. Evol.* **35**, 914–924 (2018).

PERSPECTIVE

https://doi.org/10.1038/s41586-019-1368-z

Estimating and tracking the remaining carbon budget for stringent climate targets

Joeri Rogelj[1,2,3]*, Piers M. Forster[4], Elmar Kriegler[5], Christopher J. Smith[4] & Roland Séférian[6]

Research reported during the past decade has shown that global warming is roughly proportional to the total amount of carbon dioxide released into the atmosphere. This makes it possible to estimate the remaining carbon budget: the total amount of anthropogenic carbon dioxide that can still be emitted into the atmosphere while holding the global average temperature increase to the limit set by the Paris Agreement. However, a wide range of estimates for the remaining carbon budget has been reported, reducing the effectiveness of the remaining carbon budget as a means of setting emission reduction targets that are consistent with the Paris Agreement. Here we present a framework that enables us to track estimates of the remaining carbon budget and to understand how these estimates can improve over time as scientific knowledge advances. We propose that application of this framework may help to reconcile differences between estimates of the remaining carbon budget and may provide a basis for reducing uncertainty in the range of future estimates.

Since the Fifth Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)[1], the concept of a carbon budget has risen to prominence as a tool in guiding climate policy[2]. We here define the remaining carbon budget as the finite total amount of CO_2 that can be emitted into the atmosphere by human activities while still holding global warming to a desired temperature limit. This is not to be confused with another concept, the historical carbon budget, which describes estimates of all major past and contemporary carbon fluxes in the Earth system[3]. The idea of a remaining carbon budget is grounded in well established climate science. A series of studies over the past decade has clarified and quantified why the rise in global average temperature increase is roughly proportional to the total cumulative amount of CO_2 emissions produced by human activities since the industrial revolution[4–13]. This literature has allowed scientists to define the linear relationship between warming and cumulative CO_2 emissions as the transient climate response to cumulative emissions of CO_2 (TCRE). Once established, the appeal of this concept became immediately evident: the possibility that the response of an enormously complex system—such as the response of planet Earth to our emissions of CO_2—could potentially be reduced to a roughly linear relationship would allow scientists to infer clear and easy-to-communicate implications. However, additional processes that influence and are influenced by future warming, such as the thawing of permafrost, have recently been included in models that simulate the Earth system. These additional processes add uncertainty and may change our understanding of this linear relationship. Moreover, global warming is not driven by emissions of CO_2 only. Other greenhouse gases (such as methane, fluorinated gases or nitrous oxide) and aerosols and their precursors (including soot or sulphur dioxide) affect global temperatures. Estimating the remaining carbon budget thus also implies making assumptions about these non-CO_2 contributions. This further complicates the relationship between future CO_2 emissions and global warming.

Carbon budgets nevertheless have become a powerful tool for communicating the challenges we face in aiming to hold warming to 1.5 °C and to well below 2 °C—the limits of global average temperature increase set out in the United Nations Paris Agreement[14–18]. First, every tonne of CO_2 emitted into the atmosphere by human activities adds to warming, and it does not matter whether this tonne of CO_2 is emitted today, tomorrow or yesterday. This also implies that to limit temperature increase to any level, global CO_2 emissions produced by human activities must be reduced to net-zero levels at some point in time and, on average, stay at net-zero levels thereafter. Furthermore, when aiming to limit warming to below a specific limit, a finite carbon budget also implies that the more we emit in the coming years, the faster emissions will have to decline thereafter to stay within the same budget—simple arithmetic. Finally, once net CO_2 emissions are brought to zero, warming would stabilize but would not disappear or be reversed[19–21]. Any amount by which a carbon budget compatible with a desired temperature limit is missed or exceeded would thus have to be actively and permanently removed from the atmosphere in later years. This could be achieved through measures that result in net negative CO_2 emissions, which come with their own technical and social complications[22–27]. Besides its role as a communication tool, the carbon budget concept also provides a way to exchange knowledge across disciplines. For example, such knowledge exchange is already happening for climate change mitigation requirements between the geoscience community and other disciplines that study climate change from a more societal angle[28,29].

Diversity that may confuse

Unfortunately, all that glitters is not gold. Over the past five years, a plethora of studies have been published[12,30–44] further exploring and estimating the size of carbon budgets while in some way accounting for non-CO_2 climate forcing. These studies most often focus on requirements for holding warming to the internationally agreed 1.5 °C or 2 °C limits[14–16]. Although all studies aim to evaluate the same quantity, the use of different definitions and non-CO_2 climate forcing assumptions, as well as methodological and model differences, have led to a wide variety of reported carbon budget estimates that aim to achieve temperature goals that are nominally the same (see Box 1 for an overview of carbon budget estimation approaches). This variation seems to have decreased instead of increased the broader understanding of

[1]Grantham Institute for Climate Change and the Environment, Imperial College London, London, UK. [2]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria. [3]Institute for Atmospheric and Climate Science, ETH Zurich, Zurich, Switzerland. [4]Priestley International Centre for Climate, University of Leeds, Leeds, UK. [5]Potsdam Institute for Climate Impact Research, Leibniz Association, Potsdam, Germany. [6]CNRM, Université de Toulouse, Météo-France, CNRS, Toulouse, France. *e-mail: j.rogelj@imperial.ac.uk

the remaining carbon budget and has thus tempered the initial enthusiasm about its usefulness as a guide for policy making and target setting[45,46]. This confusion is avoidable, however. Differences in remaining carbon budget estimates can be understood if a set of potential contributing factors are carefully taken into account.

Here we present a conceptual framework that allows one to track, understand, update and explain estimates of the remaining carbon budget over time. The framework's structure enables the assessment of individual contributing factors, including historical warming, the TCRE, the zero-emissions commitment and non-CO_2 contributions to future warming. It integrates suggestions made in earlier literature[12,47] and is a generalization and extension of the framework used in ref. [48].

Remaining carbon budget framework

As discussed, the remaining carbon budget can be defined as the remaining amount of CO_2 emissions that can still be emitted while keeping the global average temperature increase due to human activities to below a specific temperature limit. The framework set out below applies to a situation in which one aims to limit peak (or maximum) warming and its associated impacts. It can, however, also be extended to apply to a situation where temperature rise has temporarily exceeded an intended temperature limit, often referred to as a temperature overshoot (see Supplementary Text 1).

We present in equation (1) an estimate of the remaining carbon budget (B_{lim}) for a specific temperature change limit (T_{lim}) as a function of five terms that represent aspects of the geophysical and coupled human–environment system (equation (1): the historical human-induced warming to date (T_{hist}), the non-CO_2 contribution to future temperature rise ($T_{\mathrm{nonCO_2}}$), the zero-emissions commitment (T_{ZEC}), the TCRE, and an adjustment term for sources of unrepresented Earth system feedback (E_{Esfb}). These terms are visualized in Fig. 1 and are described and discussed in turn below.

$$B_{\mathrm{lim}} = (T_{\mathrm{lim}} - T_{\mathrm{hist}} - T_{\mathrm{nonCO_2}} - T_{\mathrm{ZEC}})/\mathrm{TCRE} - E_{\mathrm{Esfb}} \quad (1)$$

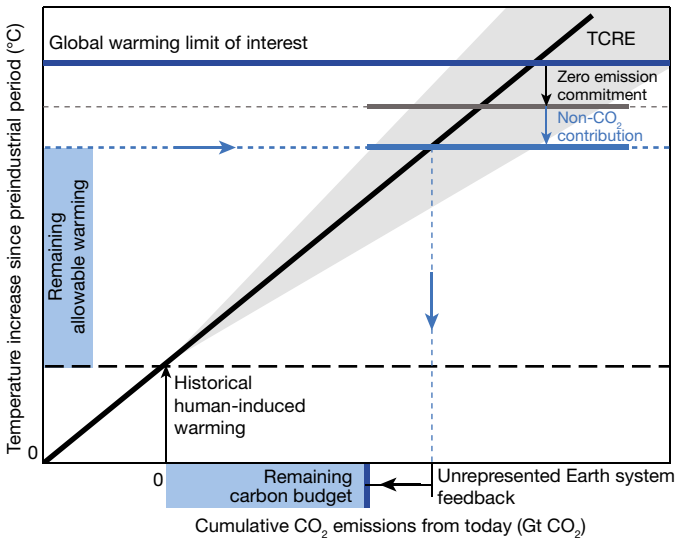


Fig. 1 | Schematic of factors contributing to the quantification of a remaining carbon budget. The schematic shows how the remaining carbon budget can be estimated from various independently assessable quantities, including the historical human-induced warming T_{hist}, the zero-emissions commitment T_{ZEC}, the contribution of future non-CO_2 warming (consistent with global net-zero CO_2 emissions or otherwise) $T_{\mathrm{nonCO_2}}$, the transient climate response to cumulative emissions of carbon (TCRE), and further correcting for unrepresented Earth system feedback E_{Esfb}. The grey shading illustrates how uncertainty in TCRE propagates from the start point. Arrows and dashed lines are visual guides illustrating how the various factors combine to provide an estimate of the remaining carbon budget. Besides estimating the remaining carbon budget B_{lim}, the framework can also be applied to understand, decompose and discuss estimates of carbon budgets calculated by other methods. The relative sizes of the various contributions shown in this schematic are not to scale.

Transient climate response to cumulative emissions

Arguably the most central term to estimating the remaining carbon budget is the TCRE (in units of °C per gigatonne of carbon dioxide (Gt CO_2); see equation (1). In essence, the remaining carbon budget is estimated by multiplying the remaining allowable warming with the inverse of the TCRE, where the magnitude of the remaining allowable warming is the result of various contributions shown in Fig. 1 and discussed below. The TCRE can be estimated from several lines of evidence, including the observational record[10,12,49–51], CO_2-only simulations[10] and multi-gas simulations[12,31,49–53] with Earth system models of varying complexity. In its latest assessment[54], the IPCC reported the TCRE to fall within the range of 0.2–0.7 °C per 1,000 Gt CO_2 with a probability of at least 66%. TCRE, and hence the linear proportionality of warming to cumulative emissions of CO_2, has also been found to be robust up to about 7,300 Gt CO_2 of cumulative emissions[54,55] and probably more[56]. This domain of application easily spans the range of carbon budgets consistent with warming limits of 1.5 °C and 2 °C.

Historical and maximum temperature increase

After TCRE, the combined remaining allowable warming (represented by $T_{\mathrm{lim}} - T_{\mathrm{hist}} - T_{\mathrm{nonCO_2}} - T_{\mathrm{ZEC}}$) is the next key determinant for estimating the remaining carbon budget. Its first term is the specific temperature limit of interest relative to preindustrial levels (T_{lim}, in units of °C), and its second term represents the historical human-induced warming (T_{hist}, in units of °C); see equation (1). T_{hist} is the amount of human-induced warming since preindustrial times until a more recent reference period, such as the 2006–2015 period.

The estimation of T_{hist} is a central factor affecting the size of the remaining carbon budget, because it determines how far we currently are from policy-relevant temperature limits (1.5 °C or 2 °C). The assessment of T_{hist} should adequately isolate the human-induced warming signal from the effects of natural forcing and variability[57,58].

The same is true for T_{lim}, and if T_{lim} is intended to represent an internationally agreed climate goal in line with the Paris Agreement it should do so by definition[15]. Two additional choices play an important role in determining or setting T_{hist} and T_{lim}: the choice of the preindustrial reference period and the temperature metric for determining global average temperature increase. Neither the preindustrial reference period nor the specific warming metric are explicitly defined by the Paris Agreement and recent literature has explored the implications and interpretations of this ambiguity[34,35,59].

The 1850–1900 period is often used as a proxy for preindustrial levels because observational temperature records stretch back to the beginning of that period[60], and key scientific reports that fed into the Paris Agreement also used this proxy[1,59,61,62] (see Supplementary Text 2 for more details). Other periods have been suggested[63–65], but ultimately the crux lies in that T_{hist} and T_{lim} should always be expressed relative to the same preindustrial reference period to avoid introducing erroneous changes to the remaining allowable warming and therewith the remaining carbon budget. Besides defining an appropriate preindustrial reference period, the choice of metric by which warming is estimated from that period is also important. Studies analysing climate model simulations or observational products can use different metrics to estimate global mean temperature change (see Supplementary Text 2). The impact of this metric choice has been highlighted recently with studies[34,59] showing that this choice can result in variations in the estimated global warming of the order of 10% (Supplementary Fig. 1), leading to a potential variation in remaining carbon budget estimates of more than 400 billion tonnes of CO_2 (ref. [59]). The IPCC has typically specified carbon budgets based on global area-averaged change in surface air temperature[48,66]. Other studies, however, have used different metrics and at times have even changed metrics between observations and projections (Supplementary Table 1, Fig. 2). This limits the comparability of these budget estimates[59]—a situation this new framework attempts to avoid.

Non-CO_2 contribution to future warming

Another term affecting the remaining allowable warming is the non-CO_2 contribution to future global temperature rise (T_{nonCO_2}, in units of °C) (see equation (1) and Fig. 1). Current and future warming depends on both CO_2-induced warming and warming due to non-CO_2 forcing. Future non-CO_2 warming might be considerable, given that reducing emissions of cooling sulphur dioxide causes warming[67] and the knowledge that no obvious mitigation options have been identified that can completely eliminate several important sources of non-CO_2 greenhouse gases[68,69]. To include T_{nonCO_2} in the remaining carbon budget framework, the non-CO_2 warming contribution between a recent reference period (for example, the same period as T_{hist}) and a specific time in the future has to be estimated. We suggest that this non-CO_2 contribution to future temperature rise should be estimated from scenarios with an internally consistent evolution of greenhouse gases and other climate forcers[36,70–74] and at the moment at which global CO_2 emissions reach net zero[48]. Estimating the non-CO_2 warming contribution at that moment in time reflects a situation in which global cumulative emissions of CO_2 are effectively capped and hence allows us to directly inform the question of how much CO_2 can be emitted while keeping warming to a given temperature level. If non-CO_2 warming were to be estimated at other moments in time, its usefulness for informing mitigation requirements would potentially be strongly reduced.

Besides the future evolution of non-CO_2 emissions, the non-CO_2 warming contribution also depends on estimates of the corresponding radiative forcing, including potential changes in surface albedo[43]. Non-CO_2 forcing and warming can be estimated with the help of simple climate models[43,75,76], inferred from more complex climate model runs[77], or taken from the literature[37,48]. Importantly, non-CO_2 emissions would continue to affect warming levels after the time when net CO_2 reaches zero, which creates uncertainty in methods that estimate budgets by integrating changes over time and after an overshoot (for example, see refs [36,43] and Box 1). These uncertainties are reduced in the framework proposed in this Perspective by focusing on the time of reaching net-zero CO_2 emissions and by considering internally consistent non-CO_2 emissions. Under these assumptions, non-CO_2 emissions are projected to result in a constant or declining forcing and warming after the time of net-zero CO_2 (refs [48,73]). However, if under alternative assumptions one would project non-CO_2 warming to continue to increase irrespective of the level of CO_2 emissions[78], this further increase should also be accounted for within T_{nonCO_2} because it would add to future peak warming.

Zero-emissions commitment

The zero-emissions commitment, T_{ZEC} (in units of °C) is the next term in the remaining carbon budget framework represented by equation (1). T_{ZEC} is defined as the additional contribution to peak warming that is still to be expected after a complete cessation of CO_2 emissions[79,80], and hence provides a correction term for the instantaneous linearity postulated by the concept of the TCRE. T_{ZEC} can be positive, negative or zero. For estimates of the remaining carbon budget, the T_{ZEC} when CO_2 emissions approach net-zero levels is of particular interest. In more general terms, this could also be formulated as an assessment of the lag in CO_2-induced warming at current and declining emissions rates[50,79]. When T_{ZEC} is positive, not all warming will have been experienced by the time global CO_2 emissions reach net zero. The estimated additional warming would hence also have to be reduced from the allowable remaining temperature increase. At present, T_{ZEC} is frequently neglected in carbon budget studies (see Supplementary Table 1, with exceptions only hypothesizing the effect of its contribution[37]) and is hence implicitly assumed to be zero or negative. Several studies suggest, however, that there might be a smaller[79–82] or larger[83,84] lag between the time when CO_2 emissions have ceased and the time of maximum warming caused by those emissions. Instead of being accounted for as a separate term, the T_{ZEC} could also be integrated within the assessment of TCRE, although a dedicated methodological framework to do so is currently lacking.

Box 1

Frequently used carbon budget definitions

Studies differ in how they define the carbon budget, and these differences affect the accuracy, size and usefulness of reported estimates. This box provides an overview of five ways in which carbon budgets can be defined, and highlights some of their strengths and weaknesses as well as how they link to the remaining carbon budget framework introduced here.

Peak or maximum temperature budgets are defined as the cumulative amount of net CO_2 emissions that would hold maximum warming to a specific temperature limit. In most cases, peak warming roughly coincides with the timing of a pathway reaching net-zero CO_2 emissions, and peak temperature budgets are thus directly compatible with the framework proposed in this paper. They also provide a direct estimate of the amount of CO_2 emissions that is consistent with achieving international temperature goals[48].

Threshold return budgets are defined as the cumulative amount of net CO_2 emissions until a specific level of warming is reached, yet only after having temporarily exceeded that level by a certain amount and during a certain period of time earlier[36,47]. By definition, they include a period of global net removal of CO_2 and hence must account for potential additional nonlinearities in the Earth system response[105]. Supplementary Text 1 clarifies how the framework presented here can be adjusted to suit this definition.

Threshold exceedance budgets are defined as the cumulative amount of net CO_2 emissions until the time at which temperature projections for a given pathway exceed a temperature threshold of interest[37]. This method has been often applied by studies that estimate carbon budgets from a limited set of simulations of complex Earth system models[10,30,32,54]. They do not provide a direct estimate of the amount of CO_2 emission that is consistent with achieving international temperature goals, but can still be discussed and understood within the framework presented in this Perspective, for example, by explicitly clarifying assumptions regarding historical warming and non-CO_2 warming at the time the temperature threshold is exceeded, and assumed values for T_{ZEC} and TCRE.

Threshold avoidance budgets are derived from emissions pathways that avoid crossing a temperature threshold of interest[37]. Their main drawback is that their definition leaves a lot of room for interpretation and variation. First, in contrast to previous budget definitions, no unambiguous point in time is available for threshold avoidance budgets until such time as net CO_2 emissions are summed, thus requiring additional assumptions[37,39]. Second, any scenario that limits warming below a threshold of interest (whether slightly or by a much larger margin) could be included in a threshold avoidance budget estimate[71]. This makes these estimates imprecise, extremely variable and difficult to compare across studies. However, even here the framework presented in this Perspective can help to structure discussions.

Finally, some studies report descriptive statistics of emissions pathways, such as cumulative CO_2 emissions until 2050 or 2100, instead of estimates of remaining carbon budgets. These statistics are not directly selected on the basis of their temperature outcome[36,71] and should not be interpreted as geophysical carbon budget requirements.

Unrepresented sources of Earth system feedback

Finally, reductions in emissions due to unrepresented Earth system feedback mechanisms (E_{Esfb}, in units of Gt CO_2), are the last term in the proposed remaining carbon budget framework (equation (1)). Any Earth system feedback that is not yet incorporated in estimates of the TCRE or that would reduce the applicability of TCRE should be assessed, and accounted for and communicated as part of E_{Esfb}.

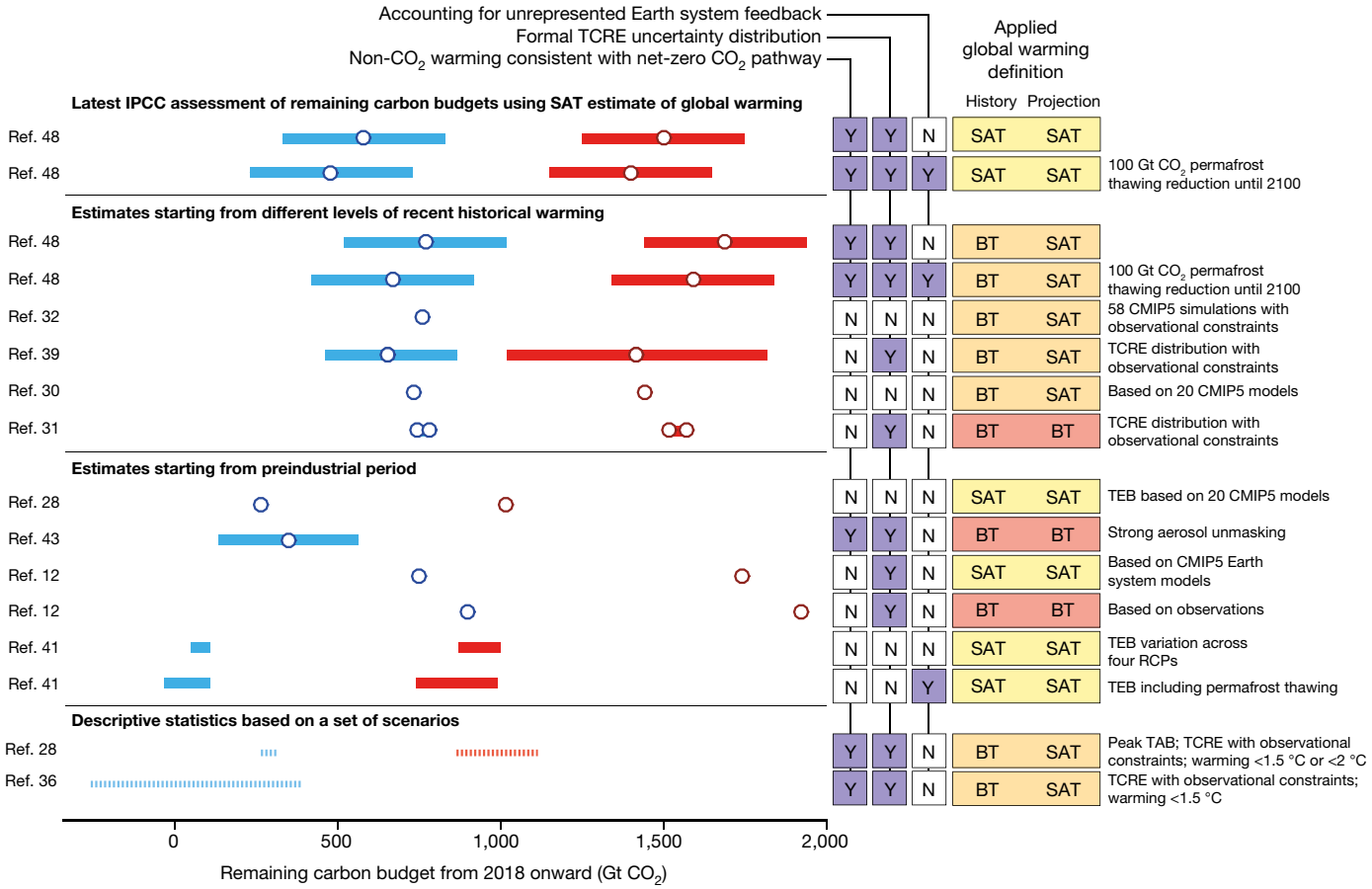


Fig. 2 | Comparison of recent remaining carbon budget estimates for limiting global warming to 1.5 °C (blue) and to 2 °C (red) relative to preindustrial levels, and overview of factors affecting their variation. Estimates are shown for a 50% probability of limiting warming to the indicated temperature levels (additional estimates for a 66% probability are provided in Supplementary Table 2). Several studies do not report formal probabilities, but report the frequency distribution across model simulations instead. The latter estimates are marked N in the 'Formal TCRE uncertainty distribution' column. Estimates shown with dashed lines indicate carbon budget estimates with an imprecise level of implied global warming, for example, because they were reported for a radiative forcing target instead. TEB means threshold exceedance budget[37]; TAB means threshold avoidance budget[37] (see Box 1). Data are taken from the IPCC Special Report on Global Warming of 1.5 °C (ref. [48]), ref. [39] (with values for 1.5 °C based on our own calculations using the same method), the IPCC Fifth Assessment Report (ref. [28]) and refs [12,30–32,36,41,43]. The latest IPCC assessment of the remaining carbon budget[48] assumes 0.97 °C of historical warming until 2006–2015, whereas other estimates assume either higher or lower warming for that period (Supplementary Table 1). The background and values for all studies are provided in Supplementary Tables 1 and 2. The assumptions made for each study are coloured (right-hand side of figure) for ease of visual grouping: N, no; Y, yes; SAT, global near-surface air temperatures; BT, blended temperatures (surface air temperature over land and sea-ice regions combined with sea surface temperature over open ocean); RCP, Representative Concentration Pathway; CMIP5, Phase 5 of the Coupled Model Intercomparison Project.

These feedback processes have typically been related to permafrost thawing[40–42,85] and the associated long-term release of CO_2 and CH_4. However, other Earth system feedback sources that can affect remaining carbon budgets have been identified[42], including changes in vegetation CO_2 uptake linked to nitrogen availability[86–88]. If unrepresented feedback results in a direct CO_2 emission from an ecosystem, the translation to the E_{Esfb} term is direct. However, because of the diverse nature of Earth system feedback[42], accounting for it through an adjustment in CO_2 emissions is not always straightforward. For example, if a feedback system results in the release of other greenhouse gases or affects the Earth system through changes in surface albedo, clouds or fire regimes, for example, its contribution needs to be translated into an equivalent CO_2 correction term (see refs [89,90] for example). Because most Earth system feedback is either sensitive to rising CO_2 or to variations in climate parameters, it is expected that these contributions are scenario-dependent, nonlinear, and in some cases realized over longer timescales only[40,41,85,91–98]. This adds to the complexity of the translation into a CO_2-equivalent correction term, and makes E_{Esfb} an uncertain contribution. E_{Esfb} could be estimated either for the time at which global net CO_2 emissions become zero or until the end of the century or beyond, assuming anthropogenic CO_2 emissions are kept at net-zero levels but feedback mechanisms continue to change over time[41,85,92,93,97]. Finally, scenario-independent Earth system feedback that scales linearly with global average temperature increase could also be incorporated by adjusting the TCRE, as long as it is not double-counted in both E_{Esfb} and TCRE.

Tracking and explaining scientific progress

We are of the opinion that through conscientious and rigorous application of the framework we propose in this Perspective, much of the confusion surrounding the size and variation of remaining carbon budget estimates can be avoided. Our proposed framework allows scientists to identify, understand and track how the progression of science on multiple fronts can affect budget estimates. It also allows us to identify and discuss key uncertainties and choices related to each respective term (Table 1). Together, these two improvements can contribute to a more constructive and informed discussion of the topic, and better communication across the various disciplines and communities that research, quantify and apply estimates of the remaining carbon budget.

The road from the geosciences to climate policy is long and winding. However, carbon budgets provide one of the simplest and most transparent means of connecting geophysical limits imposed by the Earth system to implications for climate policy. For example, they provide

Table 1 | Key choices or uncertainties of terms affecting estimates of the remaining carbon budget

Term	Symbol	Key choices or uncertainties	Type	Level of understanding
Temperature limit	T_{lim}	Choice of temperature metrics used to express global warming, choice of preindustrial reference period, and consistency with global climate goals	Choice	Medium to high
Historical human-induced warming	T_{hist}	Choice of temperature metrics used to express global warming, choice of preindustrial reference period, and consistency with global climate goals	Choice	Medium to high
Historical human-induced warming	T_{hist}	Incomplete coverage in observational data sets, and methods to estimate human-induced component	Uncertainty	Medium to high
Non-CO_2 contribution to future global warming	T_{nonCO_2}	The level of different non-CO_2 emissions that are consistent with global net-zero CO_2 emissions, which depends on policy choices but also on uncertain success of their implementation	Choice and uncertainty	Medium
Non-CO_2 contribution to future global warming	T_{nonCO_2}	Climate response to non-CO_2 forcers, particularly in the level of aerosol recovery and temperature reduction from lower methane emissions	Uncertainty	Low to medium
Zero-emissions commitment	T_{ZEC}	Sign and magnitude of zero-emission commitment at decadal time scales for current and near-zero annual CO_2 emissions	Uncertainty	Low
Transient climate response to cumulative emissions of CO_2	TCRE	Distribution of TCRE uncertainty, linearity of TCRE for increasing and stabilizing cumulative CO_2 emissions, and impact of temperature metrics on TCRE estimate	Uncertainty	Low to medium
Transient climate response to cumulative emissions of CO_2	TCRE	When extended beyond peak warming (Supplementary Text 1), uncertainty about linearity, value and distribution of TCRE for decreasing cumulative CO_2 emissions	Uncertainty	Low
Unrepresented Earth system feedback mechanisms	E_{Esfb}	Timescale and magnitude of permafrost thawing and methane release from wetlands and their representation in Earth system models, as well as other potential types of feedback	Uncertainty	Very low

Each of the terms in equation (1) is listed. 'Level of understanding' indicates our assessment of the current level of understanding of the various uncertainty components.

the geophysical foundation for setting global net-zero targets[6,99], which have recently been picked up by policy scholars as being potentially more effective in guiding policy towards a more actionable climate change mitigation goal[100]. When combined with models that simulate possible transformations to a low-carbon society[101], they can also help inform other, more specific, climate change mitigation targets.

Nevertheless, adequately characterizing and communicating the uncertainties that surround carbon budget estimates is a challenge that will remain. These uncertainties are not unfathomable, however, and language exists to describe the nature of the various uncertainty contributions[102] (Table 1, Fig. 2). In some cases, uncertainties exist because of our imprecise knowledge of certain processes or lack of precise measurements. This uncertainty is applicable to all terms in our framework except for T_{lim}, and will only gradually be reduced over time. In other cases, terms are not used consistently throughout the literature, resulting in confusion and inconsistency between carbon budget estimates (Table 1, Supplementary Table 1, Fig. 2). This is the case for the choice of global temperature metric or for the time period over which remaining carbon budgets are computed. For increased comparability and flexibility, it would be useful if global surface air temperature values were routinely estimated for observational products and if climate model projections were to report both metrics. Some uncertainties represent policy choices[44]. An example of such uncertainty is the estimate of the non-CO_2 emissions contribution to future warming. Future non-CO_2 emissions depend on future socio-economic developments and deployment of mitigation measures, and these are influenced by policy and societal choices today, for example, regarding how much the emission of non-CO_2 greenhouse gases is penalized or which sectors are targeted when promoting innovation for climate change mitigation. These policy-driven uncertainties and ambiguities can be understood, quantified and explained using a scenario-based approach. For sources of Earth system feedback that are not fully represented in models, a quantification of their impact remains difficult. Expert judgment can be applied in this case to provide an estimate of its importance.

The overview of assumptions made in carbon budget studies (Fig. 2 and Supplementary Tables 1 and 2) can already provide a first step in understanding the relative differences between estimates. For example, except for the most recent IPCC report[48], none of the estimates available in the literature simultaneously apply consistent global warming metrics for historical and projected temperatures together with a non-CO_2 warming contribution reflecting a future that is in line with the Paris Agreement (Fig. 2, Supplementary Tables 1 and 2). Several estimates also infer the chance of limiting warming to 1.5 °C from ad hoc frequency distributions of model results, instead of from a formal representation of the uncertainty in TCRE. In addition, studies typically do not include all currently identified Earth system feedback, although the impact of some has been described in dedicated studies[40–42,85].

Comparing estimates that are the same in all but their inclusion of some of the unrepresented Earth system feedback mechanisms (from refs [41,48]) suggests that the inclusion of additional Earth system feedback could consistently reduce estimates of the remaining carbon budget—something to bear in mind when future studies that use the latest generation of Earth system models become available[103]. A further insight is that estimates that apply temperature metrics other than global surface air temperatures (Fig. 2 and Supplementary Text 2) consistently suggest remaining carbon budgets that are larger than estimates that use surface air temperature only. The reasons underlying this perceived shift are well understood (see Supplementary Text 2) and can be identified as an artefact of a methodological choice. To be sure, although estimates using temperature metrics other than global averaged surface air temperature usually suggest a larger remaining carbon budget, they also come with clear climate change consequences: a relatively hotter Earth, inconsistent with the long-term temperature goal of the Paris Agreement[59]. A sound rationale therefore needs to accompany the choice of temperature metric. We strongly recommend using global average surface air temperature as the temperature metric because it is computed from invariable fields across models, model runs and over time. More detailed comparisons between published estimates of remaining carbon budgets are complicated or impossible at this stage because the quantifications of the various contributing factors by the original studies are lacking. Hence, we suggest that future studies should provide a quantitative discussion of assumptions and factors contributing to their remaining carbon budget estimates

Box 2

An application of the framework to determine the remaining carbon budget

Using the framework (equation (1)), remaining carbon budgets in line with limiting warming to 1.5 °C or to 2 °C can be estimated by drawing on information available in the literature. We now provide an example of how this could be done, starting from the assessment carried out in the context of the IPCC Special Report on Global Warming of 1.5 °C (ref. [48]).

The temperature metric is defined as follows: global warming is estimated as the global area-averaged surface air temperature change for historical warming and future projections so that T_{lim} is defined by a single consistent metric.

The preindustrial reference period is 1850–1900, as a proxy for preindustrial levels.

T_{hist} is taken to be 0.97 °C until 2006–2015 since 1850–1900. It is derived as the average over four observational datasets[60,106–110] (0.87 °C), corrected for by the ratio between surface air temperature and blended temperatures (surface air temperature over land and sea-ice regions combined with sea surface temperature over open ocean) informed by models. This level of warming is attributed to climate forcing that is caused by human activities and hence accounts for the influence of natural (internal and natural forced) variability of the climate.

T_{nonCO_2} is estimated from integrated pathways that include all climate forcing caused by human activities and derived at the time that global total CO_2 emissions reach net-zero levels[73,74]. It is estimated[75,76] to be about 0.1 °C (0–0.2 °C, 90% range) in scenarios that reach net-zero CO_2 and limit warming to 1.5 °C and to be about 0.2 °C (0.1–0.4 °C, 90% range) in scenarios limiting warming to 2 °C.

T_{ZEC} is assumed to be zero or negative, and thus not to affect the remaining allowable warming.

The remaining allowable warming starting from the recent 2006–2015 period is hence about 0.4 °C and 0.8 °C for global temperature increase limits of 1.5 °C and 2 °C, respectively.

TCRE is assumed to be normally distributed[66] with a 1σ range of 0.2–0.7 °C per 1,000 Gt CO_2.

E_{Esfb} is estimated based on literature that explicitly quantifies the effect of permafrost thawing on additional CO_2 release[40,41,85,93] and that translates the effect of other unrepresented feedback into a CO_2-equivalent correction[42]. It is estimated to reduce the remaining carbon budget by about 100 Gt CO_2 over the course of the twenty-first century, but this estimate has very low confidence attached to it (Table 1).

The combination of all terms in the framework presented here, and subtracting 290 Gt CO_2 for global CO_2 emissions since 2011, results in a remaining carbon budget B_{lim} of 480 Gt CO_2 for a 50% probability of limiting global warming to 1.5 °C (and with a B_{lim} of 740 and 320 Gt CO_2 for 33% and 66% probabilities, respectively). For a 2 °C limit, B_{lim} amounts to 1,400 Gt CO_2 for a 50% probability (and 1,930 or 1,070 Gt CO_2 for a 33% or 66% probability, respectively). In the IPCC report[48], reported numbers are 100 Gt CO_2 larger because E_{Esfb} is reported separately. In addition, the impact of varying levels of success in reducing non-CO_2 emissions can be estimated from the variation in T_{nonCO_2}, suggesting a variation of about ±250 Gt CO_2 for the remaining carbon budget for a 1.5 °C limit and −500 to +250 Gt CO_2 for a 2 °C limit.

(Supplementary Text 3), without which it is often virtually impossible to determine them later.

In the future, this framework could play a part in contextualizing new estimates, even if such estimates are obtained using alternative methods. In addition, this framework can be used in combination with expert judgment to anticipate potential changes in the remaining carbon budget. Finally, application of the framework presented here also allows us to make a more independent assessment of remaining carbon budgets by drawing on multiple lines of evidence. A simplified version of this framework has also already been applied in ref. [48] (see Box 2).

Towards more robust carbon budget estimates

Decomposing the remaining carbon budget into its contributing factors also allows one to identify a set of promising avenues for future research. An area of research that could advance this field is a closer look at TCRE. Future research is anticipated to narrow the range of best estimates of TCRE as well as to clarify the shape of the uncertainty distribution surrounding this value, the influence of a potential lag of CO_2 warming on estimating TCRE, and the validity of the TCRE concept for annual emission rates approaching net zero or during episodes of global net CO_2 removal. For example, at present there are no studies dedicated to explicit analysis of the uncertainty distribution surrounding TCRE, resulting in limited evidence to support the choice of a particular formal distribution (be it normal, lognormal, or otherwise[10,31,54]) when estimating the remaining carbon budget (see Fig. 2 and Supplementary Table 1).

Another promising area of research is the study of the interdependence between factors and their uncertainties, for example, between uncertainties in T_{hist} and T_{nonCO_2}. This could be pursued through the development of methods that allow robust estimates of recent levels of human-induced warming and allow us to link them to internally consistent projections of future non-CO_2 warming. For example, methodological developments with reduced-complexity climate models could be useful[57,75,104], because such models can flexibly and promptly incorporate most up-to-date observations and forcing estimates. This also ties into a larger question of trying to understand the overall, combined uncertainties affecting remaining carbon budgets. At present, each factor of the presented framework comes with its own uncertainties, and a method of formally combining these uncertainties is lacking.

Finally, an important uncertainty in determining the remaining carbon budget continues to be the quantification of uncertain and ill-constrained Earth system feedback processes that feed into the assessment of TCRE or E_{Esfb}. Besides affecting carbon budgets that are consistent with limiting maximum warming to a specific temperature threshold, such feedback could help to inform the risks that would be incurred by exhausting and exceeding a specific carbon budget and temperature limit, and attempting to return warming afterwards to lower levels through global net CO_2 removal (see the definition of the threshold return budget in Box 1). Challenges here lie in covering the full range of responses of these highly uncertain components, including high-risk, low-probability outcomes.

Advancements in any of these areas would improve our understanding of carbon budget estimates, and would be invaluable in the ongoing assessment of carbon budgets for the forthcoming Sixth Assessment Report of the IPCC. A systematic understanding of remaining carbon budget estimates is possible if studies improve their reporting. We recommend that future studies estimating the remaining carbon budget report the factors considered within this framework (see Supplementary Text 3 for a checklist): the surface temperature measure and historic warming used, what is assumed for TCRE, and how non-CO_2 warming and Earth system feedback sources are accounted for. A systematic understanding of remaining carbon budget estimates and how they could evolve as science advances will be essential for effective target setting and for communicating the challenges of climate change mitigation.

Online content

Any Methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available in the online version of the paper at https://doi.org/10.1038/s41586-019-1368-z.

Received: 27 October 2018; Accepted: 16 May 2019;
Published online 17 July 2019.

1. Intergovernmental Panel on Climate Change (IPCC) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, 2013).
2. Messner, D., Schellnhuber, J., Rahmstorf, S. & Klingenfeld, D. The budget approach: a framework for a global transformation toward a low-carbon economy. *J. Renew. Sustain. Energy* **2**, 031003 (2010).
3. Le Quéré, C. et al. Global carbon budget 2017. *Earth Syst. Sci. Data* **10**, 405–448 (2018).
4. Zickfeld, K., Eby, M., Matthews, H. D. & Weaver, A. J. Setting cumulative emissions targets to reduce the risk of dangerous climate change. *Proc. Natl Acad. Sci. USA* **106**, 16129–16134 (2009).
5. Matthews, H. D., Gillett, N. P., Stott, P. A. & Zickfeld, K. The proportionality of global warming to cumulative carbon emissions. *Nature* **459**, 829–832 (2009).
6. Matthews, H. D. & Caldeira, K. Stabilizing climate requires near-zero emissions. *Geophys. Res. Lett.* **35**, https://doi.org/10.1029/2007GL032388 (2008).
This was the first paper to highlight the importance of global net-zero CO_2 emissions for limiting global warming.
7. Meinshausen, M. et al. Greenhouse-gas emission targets for limiting global warming to 2 °C. *Nature* **458**, 1158–1162 (2009).
This seminal study reports carbon budgets up to the recent past and enabled the broad uptake of the carbon budget concept in climate policy discussions by linking it to the amount of carbon available in proven economically recoverable oil, gas and coal reserves.
8. Allen, M. R. et al. Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166 (2009).
9. MacDougall, A. H. & Friedlingstein, P. The origin and limits of the near proportionality between climate warming and cumulative CO_2 emissions. *J. Clim.* **28**, 4217–4230 (2015).
This paper provides a decomposition of the various factors contributing to the near-linear proportionality underlying TCRE.
10. Gillett, N. P., Arora, V. K., Matthews, D. & Allen, M. R. Constraining the ratio of global warming to cumulative CO_2 emissions using CMIP5 simulations. *J. Clim.* **26**, 6844–6858 (2013). **This study discusses the shape and observational constraints of the TCRE**.
11. Zickfeld, K. et al. Long-term climate change commitment and reversibility: an EMIC intercomparison. *J. Clim.* **26**, 5782–5809 (2013).
This multi-model study quantifies the warming commitment after a cessation of CO_2 emissions.
12. Matthews, H. D. et al. Estimating carbon budgets for ambitious climate targets. *Curr. Clim. Change Rep.* **3**, 69–77 (2017).
13. Williams, R. G., Goodwin, P., Roussenov, V. M. & Bopp, L. A framework to understand the transient climate response to emissions. *Environ. Res. Lett.* **11**, 015003 (2016).
14. The United Nations Framework Convention on Climate Change (UNFCCC) *Paris Agreement* https://unfccc.int/sites/default/files/english_paris_agreement.pdf (UNFCCC, 2015).
15. Rogelj, J., Schleussner, C.-F., & Hare, W. Getting it right matters: temperature goal interpretations in geoscience research. *Geophys. Res. Lett.* **44**, 10662–610665 (2017).
16. Schleussner, C.-F. et al. Science and policy characteristics of the Paris Agreement temperature goal. *Nat. Clim. Chang.* **6**, 827–835 (2016).
17. Knutti, R. & Rogelj, J. The legacy of our CO_2 emissions: a clash of scientific facts, politics and ethics. *Clim. Change* **133**, 361–373 (2015).
18. Matthews, H. D., Solomon, S. & Pierrehumbert, R. Cumulative carbon as a policy framework for achieving climate stabilization. *Phil. Trans. R. Soc. Lond. A* **2012**, 4365–4379 (1974).
19. Matthews, H. D. & Solomon, S. Atmosphere. Irreversible does not mean unavoidable. *Science* **340**, 438–439 (2013).
20. Solomon, S., Pierrehumbert, R., Matthews, D. & Daniel, J. in *Climate Science for Serving Society—Research, Modeling and Prediction Priorities* (eds Hurrell, J. & Asrar, G.) **506** (Springer, 2013).
21. Solomon, S. et al. Persistence of climate changes due to a range of greenhouse gases. *Proc. Natl Acad. Sci. USA* **107**, 18354–18359 (2010).
22. Minx, J. C. et al. Negative emissions—Part 1: research landscape and synthesis. *Environ. Res. Lett.* **13**, 063001 (2018).
23. Fuss, S. et al. Negative emissions—Part 2: costs, potentials and side effects. *Environ. Res. Lett.* **13**, 063002 (2018).
24. Nemet, G. F. et al. Negative emissions—Part 3: innovation and upscaling. *Environ. Res. Lett.* **13**, 063003 (2018).
25. Williamson, P. Emissions reduction: scrutinize CO_2 removal methods. *Nature* **530**, 153–155 (2016).
26. Bellamy, R. Incentivize negative emissions responsibly. *Nat. Energy* **3**, 532–534 (2018).
27. The Royal Society *Greenhouse Gas Removal* (The Royal Society, 2018).
28. Intergovernmental Panel on Climate Change (IPCC) *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (IPCC, 2014).
29. Hallegatte, S. et al. Mapping the climate change challenge. *Nat. Clim. Chang.* **6**, 663–668 (2016).
30. Millar, R. J. et al. Emission budgets and pathways consistent with limiting warming to 1.5 °C. *Nat. Geosci.* **10**, 741–747 (2017).
31. Goodwin, P. et al. Pathways to 1.5 °C and 2 °C warming based on observational and geological constraints. *Nat. Geosci.* **11**, 102–107 (2018).
32. Tokarska, K. B. & Gillett, N. P. Cumulative carbon emissions budgets consistent with 1.5 °C global warming. *Nat. Clim. Chang.* **8**, 296–299 (2018).
33. Tokarska, K. B., Gillett, N. P., Arora, V. K., Lee, W. G. & Zickfeld, K. The influence of non-CO_2 forcings on cumulative carbon emissions budgets. *Environ. Res. Lett.* **13**, 034039 (2018).
34. Richardson, M., Cowtan, K. & Millar, R. J. Global temperature definition affects achievement of long-term climate goals. *Environ. Res. Lett.* **13**, 054004 (2018).
35. Schurer, A. P. et al. Interpretations of the Paris climate target. *Nat. Geosci.* **11**, 220–221 (2018).
36. Rogelj, J. et al. Scenarios towards limiting global mean temperature increase below 1.5 °C. *Nat. Clim. Chang.* **8**, 325–332 (2018).
37. Rogelj, J. et al. Differences between carbon budget estimates unravelled. *Nat. Clim. Chang.* **6**, 245–252 (2016).
38. Rogelj, J., Meinshausen, M., Schaeffer, M., Knutti, R. & Riahi, K. Impact of short-lived non-CO_2 mitigation on carbon budgets for stabilizing global warming. *Environ. Res. Lett.* **10**, 075001 (2015).
39. Friedlingstein, P. et al. Persistent growth of CO_2 emissions and implications for reaching climate targets. *Nat. Geosci.* **7**, 709–715 (2014).
40. Comyn-Platt, E. et al. Carbon budgets for 1.5 and 2 °C targets lowered by natural wetland and permafrost feedbacks. *Nat. Geosci.* **11**, 568–573 (2018).
41. Gasser, T. et al. Path-dependent reductions in CO_2 emission budgets caused by permafrost carbon release. *Nat. Geosci.* **11**, 830–835 (2018).
This paper provides an overview of recent estimates of the impact of permafrost thawing on remaining carbon budgets.
42. Lowe, J. A. & Bernie, D. The impact of Earth system feedbacks on carbon budgets and climate response. *Phil. Trans. R. Soc. A* **376**, https://doi.org/10.1098/rsta.2017.0263 (2018).
43. Mengis, N., Partanen, A.-I., Jalbert, J. & Matthews, H. D. 1.5 °C carbon budget dependent on carbon cycle uncertainty and future non-CO_2 forcing. *Sci. Rep.* **8**, 5831 (2018).
44. Rogelj, J. et al. Mitigation choices impact carbon budget size compatible with low temperature goals. *Environ. Res. Lett.* **10**, 075003 (2015).
45. Geden, O. Politically informed advice for climate action. *Nat. Geosci.* **11**, 380–383 (2018).
46. Peters, G. P. Beyond carbon budgets. *Nat. Geosci.* **11**, 378–380 (2018).
47. Kriegler, E. et al. Pathways limiting warming to 1.5 °C: a tale of turning around in no time? *Phil. Trans. R. Soc. A* **376**, https://doi.org/10.1098/rsta.2016.0457 (2018).
48. Rogelj, J. et al. in *Global Warming of 1.5 °C: An IPCC Special Report on the Impacts of Global Warming of 1.5 °C Above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change, Sustainable Development, and Efforts to Eradicate Poverty* (eds Flato, G., Fuglestvedt, J., Mrabet, R. & Schaeffer, R.) 93–174 (IPCC/WMO, 2018).
This special report by the IPCC applied a forerunner of the framework described in this Perspective.
49. Millar, R. J. & Friedlingstein, P. The utility of the historical record for assessing the transient climate response to cumulative emissions. *Phil. Trans. R. Soc. A* **376**, https://doi.org/10.1098/rsta.2016.0449 (2018).
50. Tachiiri, K., Hajima, T. & Kawamiya, M. Increase of uncertainty in transient climate response to cumulative carbon emissions after stabilization of atmospheric CO_2 concentration. *Environ. Res. Lett.* **10**, 125018 (2015).
51. Steinacher, M. & Joos, F. Transient Earth system responses to cumulative carbon dioxide emissions: linearities, uncertainties, and probabilities in an observation-constrained model ensemble. *Biogeosciences* **13**, 1071–1103 (2016).
52. Ehlert, D., Zickfeld, K., Eby, M. & Gillett, N. The sensitivity of the proportionality between temperature change and cumulative CO_2 emissions to ocean mixing. *J. Clim.* **30**, 2921–2935 (2017).
53. MacDougall, A. H., Swart, N. C. & Knutti, R. The uncertainty in the transient climate response to cumulative CO_2 emissions arising from the uncertainty in physical climate parameters. *J. Clim.* **30**, 813–827 (2017).
54. Collins, M. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 1029–1136 (Cambridge Univ. Press, 2013).
This report by the IPCC provided the first assessment of TCRE.
55. Leduc, M., Matthews, H. D. & de Elia, R. Quantifying the limits of a linear temperature response to cumulative CO_2 emissions. *J. Clim.* **28**, 9955–9968 (2015).
56. Tokarska, K. B., Gillett, N. P., Weaver, A. J., Arora, V. K. & Eby, M. The climate response to five trillion tonnes of carbon. *Nat. Clim. Chang.* **6**, 851 (2016).
57. Haustein, K. et al. A real-time global warming index. *Sci. Rep.* **7**, 15417 (2017).
58. Huber, M. & Knutti, R. Natural variability, radiative forcing and climate response in the recent hiatus reconciled. *Nat. Geosci.* **7**, 651–656 (2014).
59. Pfleiderer, P., Schleussner, C. F., Mengel, M. & Rogelj, J. Global mean temperature indicators linked to warming levels avoiding climate risks. *Environ. Res. Lett.* **13**, 064015 (2018).
This paper quantified the impact on remaining carbon budgets of switching between global warming definitions.
60. Morice, C. P., Kennedy, J. J., Rayner, N. A. & Jones, P. D. Quantifying uncertainties in global and regional temperature change using an ensemble of observational estimates: the HadCRUT4 data set. *J. Geophys. Res. Atmospheres* **117**, https://doi.org/10.1029/2011JD017187 (2012).

61. UNFCCC *Report on the Structured Expert Dialogue on the 2013–2015 Review. FCCC/SB/2015/INF.1* http://unfccc.int/resource/docs/2015/sb/eng/inf01.pdf (UNFCCC, 2015).
62. The United Nations Environment Programme (UNEP) *The Emissions Gap Report 2014.* (UNEP, 2014).
63. Schurer, A. P., Mann, M. E., Hawkins, E., Tett, S. F. B. & Hegerl, G. C. Importance of the pre-industrial baseline for likelihood of exceeding Paris goals. *Nat. Clim. Chang.* **7**, 563–567 (2017).
64. Hawkins, E. et al. Estimating changes in global temperature since the preindustrial period. *Bull. Am. Meteorol. Soc.* **98**, 1841–1856 (2017).
65. Meinshausen, M. et al. The RCP greenhouse gas concentrations and their extensions from 1765 to 2300. *Clim. Change* **109**, 213–241 (2011).
66. Stocker, T. F. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 33–115 (Cambridge Univ. Press, 2013).
67. Samset, B. H. et al. Climate impacts from a removal of anthropogenic aerosol emissions. *Geophys. Res. Lett.* **45**, 1020–1029 (2018).
68. Smith, P. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) 811–922 (Cambridge Univ. Press, 2014).
69. Gernaat, D. E. H. J. et al. Understanding the contribution of non-carbon dioxide gases in deep mitigation scenarios. *Glob. Environ. Change* **33**, 142–153 (2015).
70. Meinshausen, M. et al. Multi-gas emission pathways to meet climate targets. *Clim. Change* **75**, 151–194 (2006).
71. Clarke, L. et al. in *Climate Change 2014: Mitigation of Climate Change. Contribution of Working Group III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Edenhofer, O. et al.) 413–510 (Cambridge Univ. Press, 2014).
72. Riahi, K. et al. The shared socioeconomic pathways and their energy, land use, and greenhouse gas emissions implications: an overview. *Glob. Environ. Change* **42**, 153–168 (2017).
73. Huppmann, D., Rogelj, J., Kriegler, E., Krey, V. & Riahi, K. A new scenario resource for integrated 1.5 °C research. *Nat. Clim. Chang.* **8**, 1027–1030 (2018).
74. Huppmann, D. et al. *IAMC 1.5 °C Scenario Explorer and Data hosted by IIASA* https://data.ene.iiasa.ac.at/iamc-1.5c-explorer/ (Integrated Assessment Modeling Consortium and International Institute for Applied Systems Analysis, 2018).
75. Smith, C. J. et al. FAIR v1.3: a simple emissions-based impulse response and carbon cycle model. *Geosci. Model Dev.* **11**, 2273–2297 (2018).
76. Meinshausen, M., Raper, S. C. B. & Wigley, T. M. L. Emulating coupled atmosphere-ocean and carbon cycle models with a simpler model, MAGICC6—Part 1: model description and calibration. *Atmos. Chem. Phys.* **11**, 1417–1456 (2011).
77. Myhre, G. et al. in *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (eds Stocker, T. F. et al.) 659–740 (Cambridge Univ. Press, 2013).
78. Kriegler, E. et al. Fossil-fueled development (SSP5): an energy and resource intensive scenario for the 21st century. *Glob. Environ. Change* **42**, 297–315 (2017).
79. Ehlert, D. & Zickfeld, K. What determines the warming commitment after cessation of CO_2 emissions? *Environ. Res. Lett.* **12**, 015002 (2017).
80. Gillett, N. P., Arora, V. K., Zickfeld, K., Marshall, S. J. & Merryfield, W. J. Ongoing climate change following a complete cessation of carbon dioxide emissions. *Nat. Geosci.* **4**, 83–87 (2011).
81. Ricke, K. L. & Caldeira, K. Maximum warming occurs about one decade after a carbon dioxide emission. *Environ. Res. Lett.* **9**, 124002 (2014).
82. Zickfeld, K. & Herrington, T. The time lag between a carbon dioxide emission and maximum warming increases with the size of the emission. *Environ. Res. Lett.* **10**, 031001 (2015).
83. Frölicher, T. L. & Paynter, D. J. Extending the relationship between global warming and cumulative carbon emissions to multi-millennial timescales. *Environ. Res. Lett.* **10**, 075002 (2015).
84. Frölicher, T. L., Winton, M. & Sarmiento, J. L. Continued global warming after CO_2 emissions stoppage. *Nat. Clim. Chang.* **4**, 40–44 (2014).
85. MacDougall, A. H., Zickfeld, K., Knutti, R. & Matthews, H. D. Sensitivity of carbon budgets to permafrost carbon feedbacks and non-CO_2 forcings. *Environ. Res. Lett.* **10**, 125003 (2015).
86. Zaehle, S. et al. Evaluation of 11 terrestrial carbon–nitrogen cycle models against observations from two temperate free-air CO_2 enrichment studies. *New Phytol.* **202**, 803–822 (2014).
87. Wenzel, S., Cox, P. M., Eyring, V. & Friedlingstein, P. Projected land photosynthesis constrained by changes in the seasonal cycle of atmospheric CO_2. *Nature* **538**, 499–501 (2016).
88. Arneth, A. et al. Terrestrial biogeochemical feedbacks in the climate system. *Nat. Geosci.* **3**, 525–532 (2010).
This review presents an overview of terrestrial Earth system feedback mechanisms that could further affect TCRE and estimates of remaining carbon budgets.
89. Carrer, D., Pique, G., Ferlicoq, M., Ceamanos, X. & Ceschia, E. What is the potential of cropland albedo management in the fight against global warming? A case study based on the use of cover crops. *Environ. Res. Lett.* **13**, 044030 (2018).
90. Allen, M. R. et al. A solution to the misrepresentations of CO_2-equivalent emissions of short-lived climate pollutants under ambitious mitigation. *npj Clim. Atmos. Sci.* **1**, 16 (2018).
91. Burke, E. J. et al. Quantifying uncertainties of permafrost carbon–climate feedbacks. *Biogeosciences* **14**, 3051–3066 (2017).
92. Schneider von Deimling, T. et al. Observation-based modelling of permafrost carbon fluxes with accounting for deep carbon deposits and thermokarst activity. *Biogeosciences* **12**, 3469–3488 (2015).
93. Schneider von Deimling, T. et al. Estimating the near-surface permafrost-carbon feedback on global warming. *Biogeosciences* **9**, 649–665 (2012).
94. Schuur, E. A. G. et al. Climate change and the permafrost carbon feedback. *Nature* **520**, 171–179 (2015).
95. Schaefer, K., Lantuit, H., Romanovsky, V. E., Schuur, E. A. G. & Ronald Witt, R. The impact of the permafrost carbon feedback on global climate. *Environ. Res. Lett.* **9**, 085003 (2014).
96. Koven, C. D. et al. A simplified, data-constrained approach to estimate the permafrost carbon–climate feedback. *Phil. Trans. R. Soc. A* **373**, https://doi.org/10.1098/rsta.2014.0423 (2015).
97. MacDougall, A. H. & Knutti, R. Projecting the release of carbon from permafrost soils using a perturbed parameter ensemble modelling approach. *Biogeosciences* **13**, 2123–2136 (2016).
98. Schwinger, J. & Tjiputra, J. Ocean carbon cycle feedbacks under negative emissions. *Geophys. Res. Lett.* **45**, 5062–5070 (2018).
99. Rogelj, J. et al. Zero emission targets as long-term global goals for climate protection. *Environ. Res. Lett.* **10**, 105007 (2015).
100. Geden, O. An actionable climate target. *Nat. Geosci.* **9**, 340 (2016).
101. Weyant, J. Some contributions of integrated assessment models of global climate change. *Rev. Environ. Econ. Policy* **11**, 115–137 (2017).
102. Smith, L. A. & Stern, N. Uncertainty in science and its role in climate policy. *Phil. Trans. R. Soc. A* **369**, 4818–4841 (2011).
103. Eyring, V. et al. Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016).
104. Meinshausen, M., Wigley, T. M. L. & Raper, S. C. B. Emulating atmosphere–ocean and carbon cycle models with a simpler model, MAGICC6—Part 2: Applications. *Atmos. Chem. Phys.* **11**, 1457–1471 (2011).
105. Zickfeld, K., MacDougall, A. H. & Matthews, H. D. On the proportionality between global temperature change and cumulative CO_2 emissions during periods of net negative CO_2 emissions. *Environ. Res. Lett.* **11**, 055006 (2016).
106. Allen, M. R. et al. Framing and context. In *Global Warming of 1.5 °C. An IPCC Special Report on the Impacts of Global Warming of 1.5 °C Above Pre-Industrial Levels and Related Global Greenhouse Gas Emission Pathways, in the Context of Strengthening the Global Response to the Threat of Climate Change* (eds Masson-Delmotte, V. et al.) 47–92 (IPCC/WMO, 2018).
107. Cowtan, K. & Way, R. G. Coverage bias in the HadCRUT4 temperature series and its impact on recent temperature trends. *Q. J. R. Meteorol. Soc.* **140**, 1935–1944 (2014).
108. Vose, R. S. et al. NOAA's merged land–ocean surface temperature analysis. *Bull. Am. Meteorol. Soc.* **93**, 1677–1685 (2012).
109. Karl, T. R. et al. Possible artifacts of data biases in the recent global surface warming hiatus. *Science* **348**, 1469–1472 (2015).
110. Hansen, J., Ruedy, R., Sato, M. & Lo, K. Global surface temperature change. *Rev. Geophys.* **48**, RG4004 (2010).

Acknowledgements We acknowledge support from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement number 641816 (CRESCENDO) and grant agreement number 820829 (CONSTRAIN), and from the UK Natural Environment Research Council (NERC) under project NE/N006038/1 (SMURPHS). We thank J. Cook and colleagues contributing to the IPCC Special Report on Global Warming of 1.5 °C (ref. [48]) for comments.

Reviewer information *Nature* thanks Edward Parson, Nathan Gillett and Pierre Friedlingstein for their contribution to the peer review of this work.

Author contributions J.R. coordinated the paper. All authors contributed substantially to the development of the framework, its description and presentation, and the writing of the paper. C.J.S. produced Supplementary Fig. 1. J.R. carried out the comparison of remaining carbon budgets, produced Figs. 1 and 2, and led the writing of the paper.

Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1368-z.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to J.R.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

ARTICLE

https://doi.org/10.1038/s41586-019-1384-z

Holistic prediction of enantioselectivity in asymmetric catalysis

Jolene P. Reid[1] & Matthew S. Sigman[1]*

When faced with unfamiliar reaction space, synthetic chemists typically apply the reported conditions (reagents, catalyst, solvent and additives) of a successful reaction to a desired, closely related reaction using a new substrate type. Unfortunately, this approach often fails owing to subtle differences in reaction requirements. Consequently, an important goal in synthetic chemistry is the ability to transfer chemical observations quantitatively from one reaction to another. Here we present a holistic, data-driven workflow for deriving statistical models of one set of reactions that can be used to predict out-of-sample reactions. As a validating case study, we combined published enantioselectivity datasets that employ 1,1′-bi-2-naphthol (BINOL)-derived chiral phosphoric acids for a range of nucleophilic addition reactions to imines and developed statistical models. These models reveal the general interactions that impart asymmetric induction and allow the quantitative transfer of this information to new reaction components. This technique creates opportunities for translating comprehensive reaction analysis to diverse chemical space, streamlining both catalyst and reaction development.

The efficacy of a catalytic process is dictated by the possible transition states, which feature core non-covalent interactions that determine their geometries and energies[1,2]. Such interactions are often difficult to identify and define because they are energetically weak and sensitive to the molecular properties of every reaction component (catalyst, substrates, reagents, solvent and so on)[3,4]. This overarching issue in reaction optimization is often exacerbated by subtle connections across several reaction variables, wherein modest structural changes to any or a few of these can have a profound effect on the experimental outcome[5–7]. These factors, combined with the number of dimensions under study in most reactions, are the underlying reasons that optimization is traditionally empirical[8,9]. This situation is particularly common in the area of asymmetric catalysis, wherein seemingly minor structural variations in any reaction component can have acute and non-intuitive influences on the observed enantioselectivity[10]. However, it is possible that such mechanistic outliers may be concealed within larger datasets because our pattern recognition skills do not perceive pivotal generalities when reaction situations change.

On this basis, we hypothesized that connecting common mechanistic features through the simultaneous interrogation of all reaction components would provide a holistic view of the key non-covalent interactions responsible for reaction performance. This would enable the transfer of experimental observations to genuinely different substrate combinations with unique catalysts. Here we develop and deploy a workflow that parameterizes all the reaction variables of more than 350 distinct reaction combinations, which allows the development of comprehensive statistical models, resulting in the ability to predict reaction performance for entirely different structural motifs. The workflow includes techniques to probe general mechanistic principles, which provides the basis for transfer learning or generalized identification of the key interactions imparting asymmetric induction.

Asymmetric catalysis is replete with examples of catalysts that can promote disparate reactions through a common mode of activation[11–14]. However, when 'similar' reactions are attempted, many changes to the precise reaction conditions are often required to obtain the desired reaction performance[15,16]. These changes can be subtle (that is, one aromatic solvent for another) or more profound (one catalyst class for another). This leads us to ask (1) whether mechanistic insight is transferable to a new reaction in the same subclass, given that a standard mechanistic paradigm may exist with a general mode of activation? If so, (2) how could a data-driven workflow that combines data acquisition and a description of the molecules involved mathematically be used to build a statistical model for diverse and multiple reaction profiles? And if such a workflow is achievable, (3) can the observed conditions of one or more reactions be deployed to predict the performance of another? Such analysis could provide a mechanistic understanding of why certain conditions are effective for a general reaction type and the ability to transfer this information quantitatively to out-of-sample predictions streamlining reaction optimization[17,18].

To assess a specific workflow that is designed to probe the questions posed above, it would be pragmatic to compare transformations within a reaction class facilitated by a single catalyst chemotype. Although multifarious reports of the same catalyst class for different transformations exist in enantioselective catalysis, comparative studies—even qualitative rather than quantitative—have been sparse. Such an assessment would be challenging because most datasets, often generated under non-uniform conditions, are incomplete and readily comprehensible descriptors for each varying reaction component need to be developed. To address this correlation challenge, we envisioned a strategy for the interrogation of enantioselective catalysis involving the application of modern data-analysis methods and advanced parameter sets. In this approach, integrated descriptor sets—quantitative structure–activity relationships (QSAR), molecular mechanics (MM) and density functional theory (DFT) derived)[19]—are related to a relatively large library of outputs collected from a general reaction and catalyst type, which are data-mined from multiple literature sources (see the Supplementary Information). By combining appropriate data-organization and trend-analysis techniques, general relationships between reactions can be established. The ability of the statistical models to predict a new reaction type performance is used as a validation of mechanistic transferability (Fig. 1).

[1]Department of Chemistry, University of Utah, Salt Lake City, UT, USA. *e-mail: sigman@chem.utah.edu

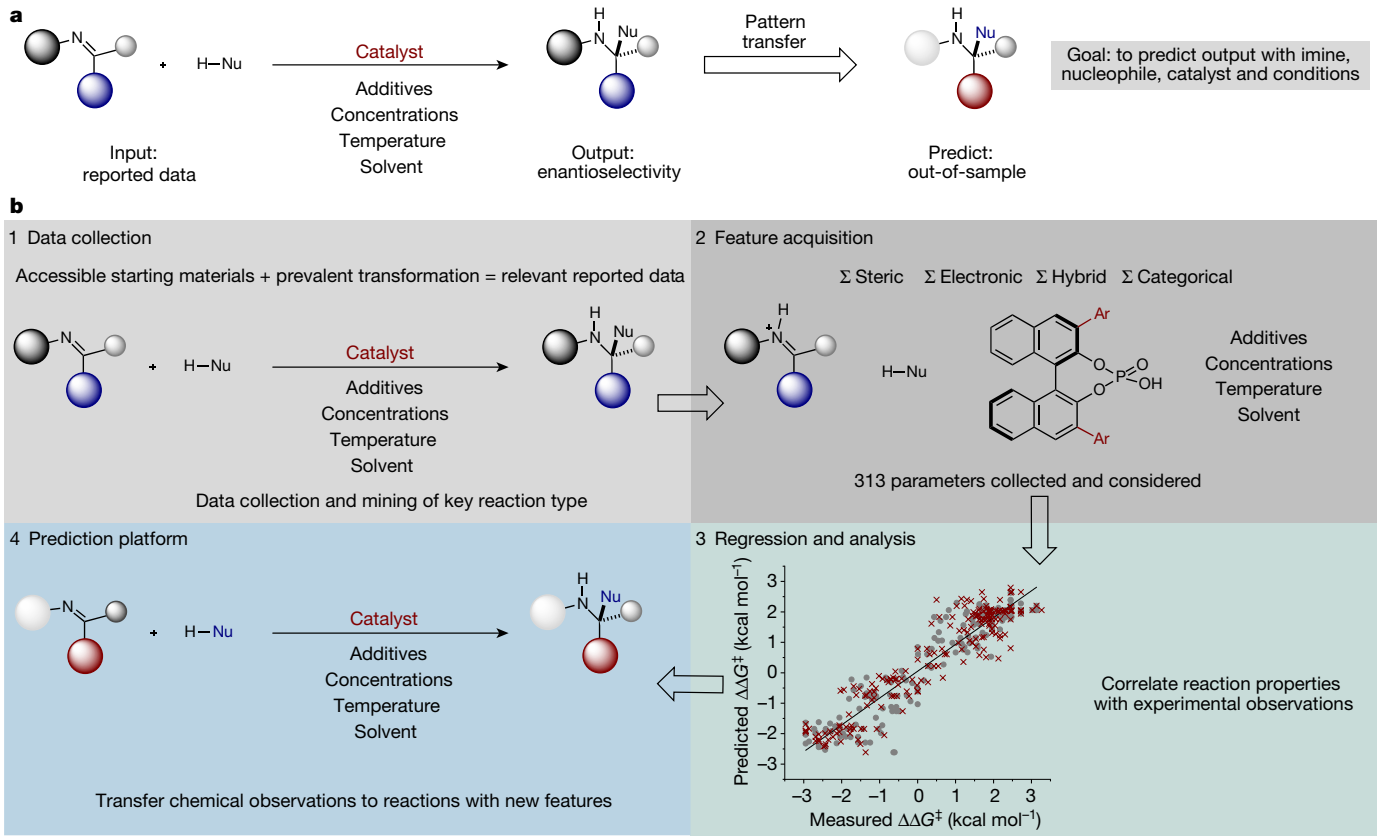


Fig. 1 | Workflow for interrogating and applying mechanistic transferability. a, Mechanistic transferability. BINOL-based phosphoric acid catalysed nucleophilic additions to imines as a general reaction for workflow development. **b**, Prediction workflow. Reaction performance predictions are streamlined by employing a mechanistic transferability strategy implemented by correlating all reaction variables to enantioselectivity. General correlations can be built to reveal the interactions between any reaction component in the relevant transition state and enantioselectivity. The mechanistic principles leading to enantioselective catalysis captured by the statistical models can be transferred to genuinely different structural motifs not contained in the training dataset. Σ indicates the totality of the descriptor categories that were considered.

Reaction platform selection

As a proof-of-concept reaction class, we chose the addition of various nucleophiles to imines owing to the ubiquity of this type of transformation in asymmetric catalysis[20,21]. This reaction class uses imine starting materials that are easy to obtain and the resulting amine products have broad applicability in both synthetic and biosynthetic settings[22,23]. As a next step, we evaluated the different catalyst chemotypes used in this reaction class, focusing on those that provide a wide range of both substrate structural types and enantioselectivity data from published sources. With these constraints in mind, we selected the field of chiral phosphoric acid (CPA) catalysis, in particular the addition of protic nucleophiles to imines catalysed by chiral 1,1′-bi-2-naphthol (BINOL)-derived phosphoric acids bearing aromatic groups at the 3 and 3′ positions (Fig. 1)[24].

To initiate this workflow, an expanded inventory of 367 reactions with varied components was curated from multiple reports (for a list of references, see Supplementary Information). From this survey, we categorized the dataset by imine transition-state geometry (*E* or *Z*) wherein *E*-imine transition states have a +e.e. value and *Z*-imines have a −e.e. value. Imine stereochemistry was determined by the enantiomer of the product formed if the imine was derived from an aldehyde. However, if ketimines (imines derived from ketones) were employed, we also needed to consider substituent size if the smaller C-substituent has higher Cahn–Ingold–Prelog (CIP) priority[25,26]. For the reactions we studied here, this affects only ketimines that have either a trifluoromethyl or ester C-substituent, which are considered to have lower priority for the purpose of assigning an *E* or *Z* transition state. This is important in understanding product enantioselectivities, because nucleophile addition to the same face will yield opposite enantiomers for the *E* and *Z* configurations. Therefore, the models developed will not be capable of predicting product stereochemistry but can be deployed to predict whether a reaction will proceed via an *E*- or *Z*-type mechanism and this information can be used to determine absolute configuration.

Simultaneously, we collected a diverse array of molecular descriptor values from DFT-optimized geometries to describe the structural features of each imine, nucleophile, catalyst and solvent. Unfortunately, the lack of structural commonality for particular molecular subsets creates a challenge in identifying readily comprehensible and extensive parameter sets for each component. For example, when comparing substrates and catalyst structures, it is apparent that they have overlapping and distinctive features that are probably required for determining selectivity patterns (Extended Data Fig. 1). By contrast, the solvents do not have common substructures, yet are critical for enantioselectivity.

To address this limitation, we explored two approaches: (1) we collected parameters derived from DFT calculations, which satisfactorily describe molecules containing common structural features including Sterimol parameters, bond lengths, angle measurements, molecular vibrations and intensities, natural bond orbital (NBO) charges, polarizabilities, highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies[27,28]. We collected these parameters for both the reaction partners and the catalysts. (2) We used two-dimensional descriptors (such as topology and connectivity as exemplified by molecular shape, size and number of heteroatoms) because this is a traditional method of assessing structurally disparate molecules such as solvents[29,30]. Other reaction variables, such as concentration of reagents or catalysts and inclusion of molecular sieves, were also included as categorical descriptors (see Supplementary Information).

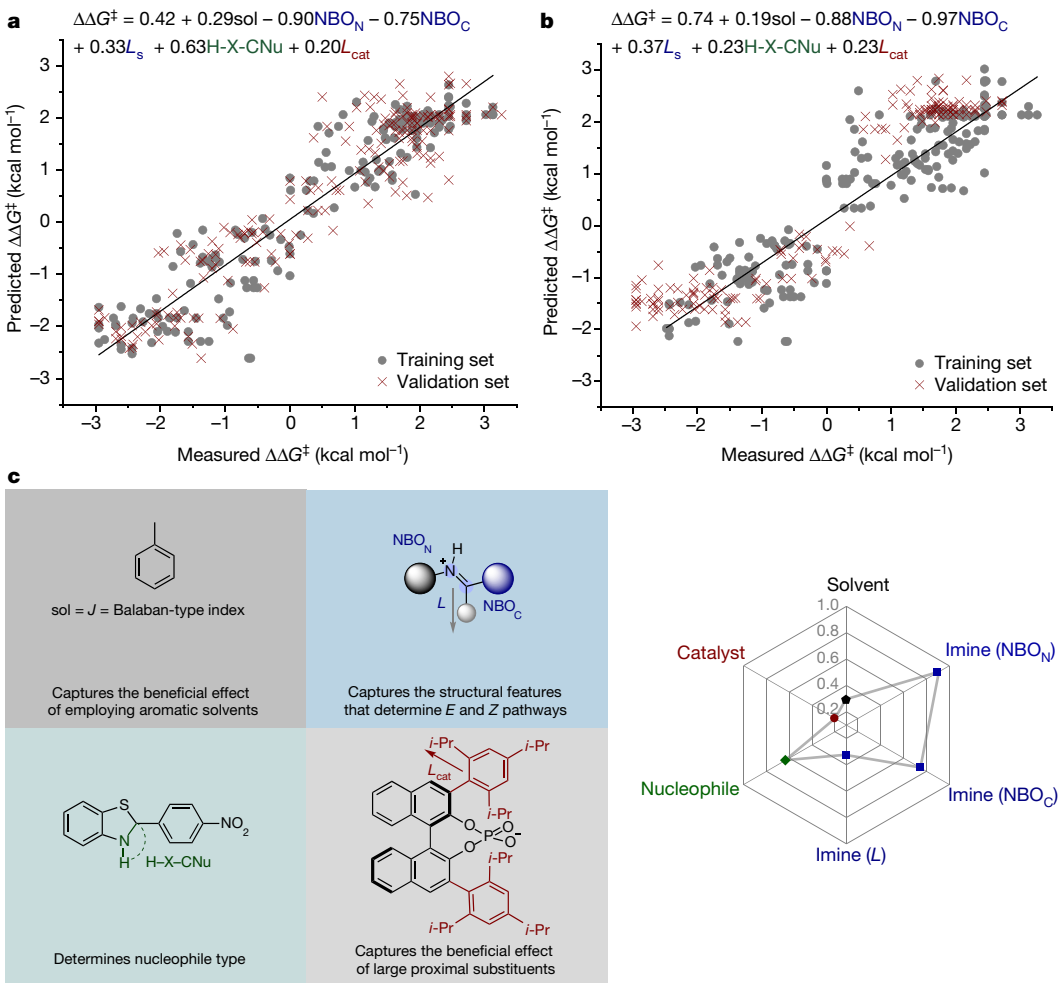


Fig. 2 | Comprehensive model development. a, Comprehensive regression model containing 367 data entries facilitated by parameterization of every reaction variable. 'sol' is the solvent term, 'NBO$_N$' and 'NBO$_C$' are imine natural bond orbital parameters, L_s is a steric descriptor of the smallest imine substituent, 'H–X–CNu' is the nucleophile angle measurement and L_{cat} is the length of the catalyst 2-substituent. A positive percentage enantiomeric excess (% e.e.) value indicates the *E*-imine transition state, and a negative percentage enantiomeric excess value indicates the *Z*-imine transition state. The line is a fit, $y = 0.88x + 0.05$. The leave-one-out (LOO) cross-validation score is 0.87; the average *k*-fold (here, fourfold) cross-validation score is 0.87; the goodness of fit R^2 is 0.88; the predicted R^2 is 0.87. **b**, Test of mechanistic transferability in the dataset via leave-one-reaction-out (LORO) analysis. Distinct reactions (as determined by individual publications) are defined as the validation set. The line is a fit, $y = 0.84x + 0.12$. R^2 is 0.84; the R^2 predicted using LORO (here, seven reactions were left out) is 0.85. **c**, Visual analysis and interpretation of the model terms (coefficients are shown).

Comprehensive model development

Linear regression algorithms (see Supplementary Information) were then applied to the entire dataset (367 reactions) to identify correlations between the molecular structure of every reaction variable defined by the parameters collected in the previous step of the workflow and the experimentally determined enantioselectivity. $\Delta\Delta G^{\ddagger} = -RT\ln(\text{e.r.})$ (where e.r. is the enantiomeric ratio, T is the temperature at which the reaction was performed and R is the gas constant) was regressed to an equation to reveal a surprisingly good correlation despite the large structural variance included in the training set. Both cross-validation analysis (leave-one-out (LOO) and *k*-fold) and external validation, in which the dataset is partitioned pseudorandomly into 50:50 training:validation sets, suggest a relatively robust model (see Supplementary Information). The model emphasizes solvent (black), imine (blue), nucleophile (green) and catalyst (red) terms distributed over six parameters, as contributors to the enantioselectivity across these seventeen reaction types (Fig. 2a). A slope approaching unity and intercept approaching zero over the training set indicates an accurate and predictive model with a goodness-of-fit R^2 value of 0.88, demonstrating a high degree of precision. The largest coefficients in this normalized model belong to the imine NBO descriptors, indicating the crucial role of the imine substrate in the quantification of enantioselectivity as highlighted by the formation of both enantiomeric products, a consequence of active *E* and *Z* configurations (see below). A comparison of two Strecker reactions performed under uniform conditions results in values ranging from +99% enantiomeric excess for the enantiomer that proceeds through the *E*-imine transition state and −80% enantiomeric excess for the *Z*-imine transition state. Remarkably, this represents a 3.5 kcal mol^{-1} energy range, based solely on imine structure.

We postulated that the ability to correlate and predict using a singular model for an array of reactions suggests that the transition-state features are fundamentally similar within this reaction range. Perhaps the best test of this hypothesis could be achieved by a 'leave one reaction out' (LORO) analysis. In this statistical evaluation, the catalyst, imine and nucleophile structures are varied as a validation set and assessed through the ability of the model to predict with sufficient accuracy. This would report on the model's capacity to match patterns across a general reaction type. Using this analysis, each distinct reaction (as determined by individual publications) in the data field was evaluated, with most predicted well (see Supplementary Information). As an illustration of model robustness, we could exclude up to seven reactions with little change in the correlation statistics (Fig. 2b). However, not surprisingly, some reactions were poorly predicted using the LORO protocol, which can be attributed to the model's inability to capture specific structure changes if they are not adequately expressed in the training set. In sum, the descriptor definitions coupled to the model and validation strategies do demonstrate that patterns can be matched. This is consistent with the hypothesis that a defined set of key non-covalent interactions impart asymmetric induction across a general reaction type. Essentially, this workflow provides evidence that one reaction can be used to predict the results of another, quantitatively.

Trend analysis

Although the comprehensive model in Fig. 2 establishes the capacity of the selected parameters to describe general aspects of this system, the ultimate goal of our workflow is to discern subtle underlying mechanistic phenomena. This objective could not be achieved by using the above correlation because it was produced by using the entire dataset, which provides only an overview of the mechanistic patterns.

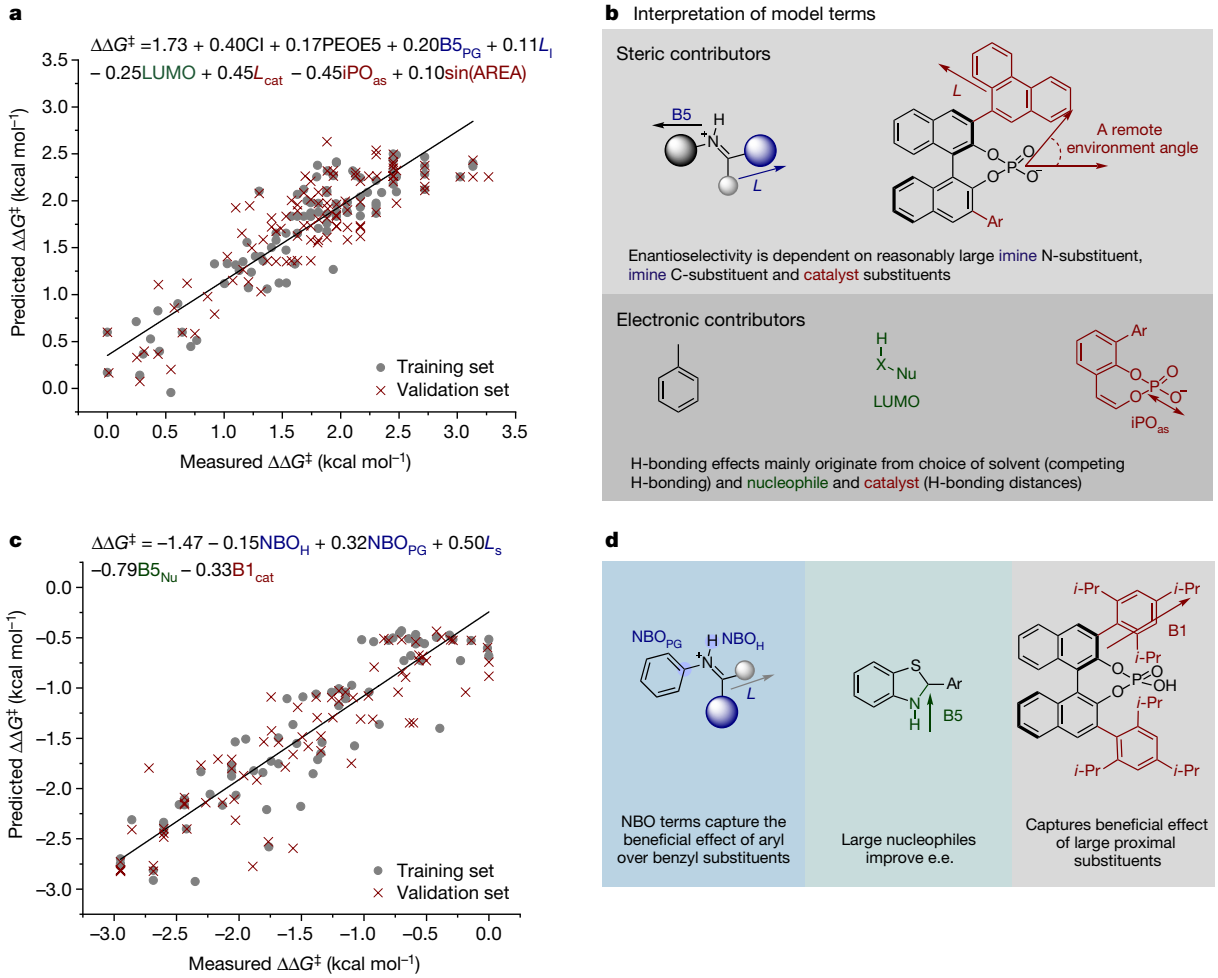


Fig. 3 | Development of focused correlations. a, Regression *E*-imine model containing 204 entries data-mined from nine literature sources (see the Supplementary Information for references). 'CI' and 'PEOE5' are solvent descriptors, '$B5_{PG}$' and L_I are the imine steric descriptors, LUMO is the lowest unoccupied molecular orbital energy describing the nucleophile, L_{cat} is the length of the catalyst 2-substituent, 'iPO_{as}' is the P–O asymmetric stretching intensity and 'AREA' is a remote environment angle. The line is a fit, $y = 0.80x + 0.35$. The LOO cross-validation score is 0.76; the average *k*-fold (here, fourfold) cross-validation score is 0.74; R^2 is 0.80; the predicted R^2 is 0.73. **b**, Interpretation of *E*-imine model terms. The model emphasizes the importance of both steric and electronic factors. Reasonably large catalyst and imine substituents lead to high levels of enantioselectivity; if these two components are matched any nucleophile should be compatible. **c**, Regression *Z*-imine model containing 147 entries data-mined from eight literature sources (see the Supplementary Information for references). 'NBO_H' and 'NBO_{PG}' are the imine natural bond orbital parameters; L_s is a steric descriptor of the smallest imine substituent; '$B5_{Nu}$' is the nucleophile steric descriptor and '$B1_{cat}$' is the Sterimol B1 term. The line is a fit, $y = 0.83x - 0.24$. The LOO cross-validation score is 0.80; the average *k*-fold (here, fourfold) cross-validation score is 0.79; R^2 is 0.83; the predicted R^2 is 0.80. **d**, Interpretation of *Z*-imine model terms. Overlapping steric terms describing the catalyst and the imine reinforce the notion that similar interactions remain within the two geometric imine stereoisomers. However, this model emphasizes the importance of steric contributions predominantly from the nucleophile for high enantioselectivities.

We hypothesized that a series of focused correlations, coupled with an evaluation of the overall trends, might serve to reveal fundamental features of the systems. To this end, we truncated the dataset into subsets, categorized by imine transition-state geometry (*E* or *Z*) determined by the relative sign of the enantiomeric excess defined previously, as these are hypothesized to lead to structurally distinct interactions with the other reaction components. This organizational scheme was viewed as a means of facilitating the identification of catalyst features that affect particular mechanistic pathways and therefore, reactant combinations (and vice versa). Linear regression algorithms were then applied to this data classification to identify correlations between molecular structure and the experimentally determined enantioselectivity. Subsequently, analysis and refinement of the resulting models were used to produce explicit mechanistic hypotheses (Fig. 3).

The correlation depicted in Fig. 3 was identified from a set of 204 reactions (evenly split into training and validation sets) that proceed via the *E*-imine transition state. The relationship includes two solvent, two imine, one nucleophile and three catalyst terms. Overall, the statistical model suggests a mechanistic scenario in which the imine adopts an arrangement that minimizes energetically penalizing repulsion interactions with reasonably large catalyst substituents[31]. Perhaps most telling is that the steric profile of the nucleophile does not have much effect on the stereoselectivity outcome, despite the large structural variance. The included parameters (LUMO and the P–O asymmetric stretching intensity, iPO_{as}) suggest that hydrogen-bonding contacts between catalyst and nucleophile play a minor part and the use of almost any nucleophile should be compatible with the reaction if the imine and catalyst are matched.

In evaluating the model for *Z*-imines determined by 147 reactions, a number of overlapping terms reinforce the notion that similar interactions between catalyst and substrates remain within the two geometric imine stereoisomers. Two of these terms—the size of the catalyst aryl substituent as measured by the Sterimol B1 term and the imine NBO parameter—essentially describe the repulsive interactions between proximal sterics and the imine *N*-substituent, a critical catalyst–substrate interaction common to both transition-state imine

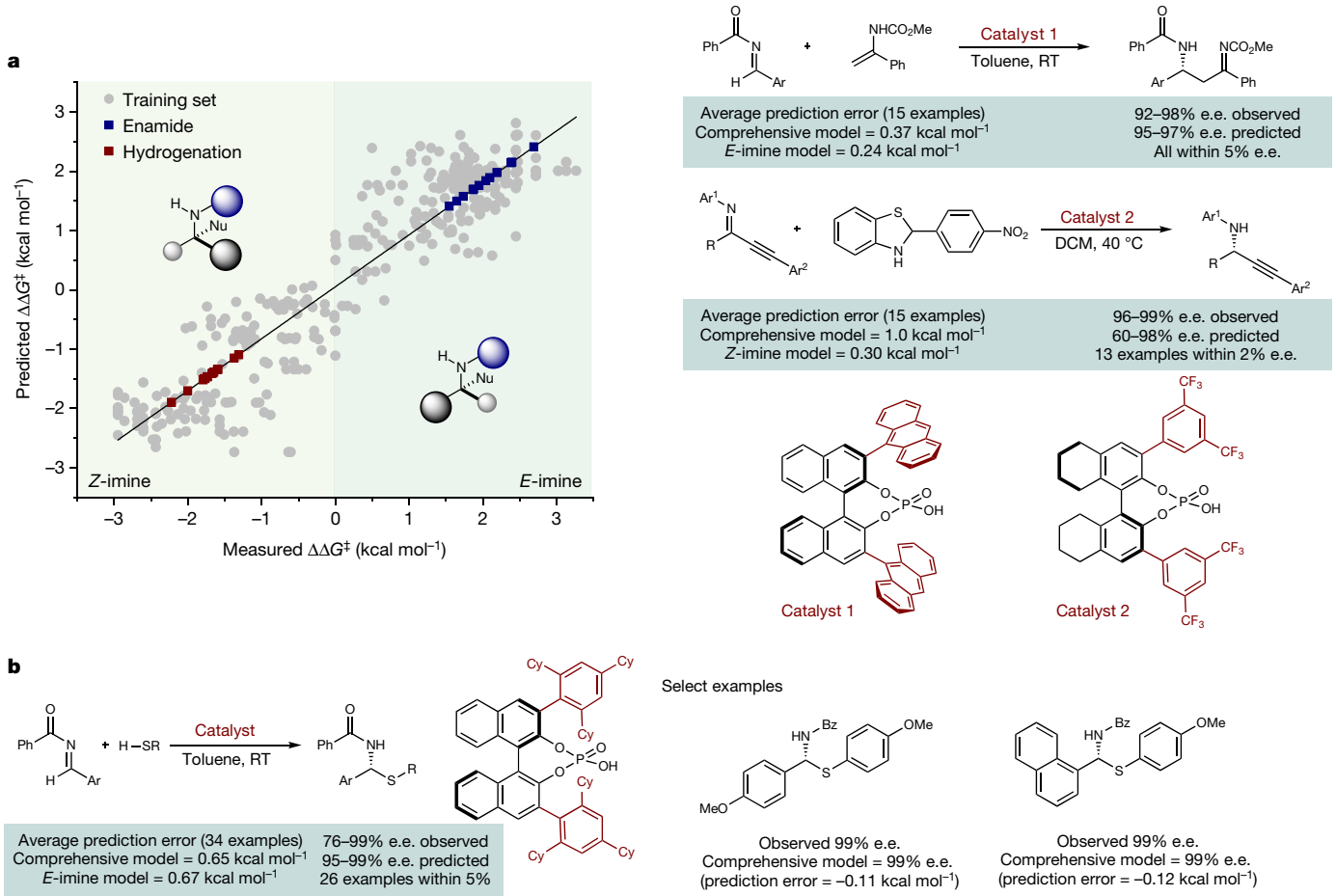


Fig. 4 | Out-of-sample predictions using two-tiered prediction workflow. Comprehensive model first determines the *E* or *Z* transition state, configuration specific models are then used to refine predictions. A generic amine product denotes the stereochemical outcome predicted if the reaction proceeds via the *E* or *Z* transition state and is catalysed by an (*R*)-CPA. Product stereochemistry is reversed if (*S*)-CPA is used. **a**, Out-of-sample prediction. Application to addition of enecarbamates to benzoyl imines and transfer hydrogenation of alkynyl ketimines. DCM, dichloromethane; RT, room temperature (25 °C). **b**, Out-of-sample prediction and extrapolation. Prediction of TCYP, which has cyclohexyl groups at the 2,4,6 positions of the aromatic ring, to be a highly selective catalyst for the addition of thiol to benzoyl imines.

configurations. The most compelling difference between the two models is that the *Z*-imine model includes an important nucleophile steric descriptor, which is the most highly weighted term in the equation. This suggests that larger nucleophiles introduce enhanced repulsive interactions with the catalyst substituents in the transition state, leading to the competing product, which ultimately favours the observed enantiomer. This claim is further supported by the observation of high enantioselectivities when using catalysts with smaller substituents (for example, Ar = 3,5-$(CF_3)_2C_6H_3$). The proposed physical meanings of each term in the mathematical equations have been summarized in Fig. 3.

Evaluation of prediction capabilities

As a final step in the workflow, we evaluated the ability to transfer the mechanistic principles leading to enantioselective catalysis captured by the statistical models to genuinely different structural motifs not contained in the training dataset. If effective out-of-sample prediction were possible, the model could predict the impact of a new imine, nucleophile and/or catalyst. Initially, reaction performance was evaluated using the comprehensive model to determine the mechanistic pathway under operation, and these predictions could then be further refined with the specific models (*E* or *Z*). This two-tiered workflow is imperative because the process avoids mechanistic assumptions about whether the reaction proceeds via an *E* or *Z* transition state, thus ensuring that the results of the test reactions are unknown. The comprehensive model does not immediately allow prediction of stereochemistry; however, product configuration can be assigned from the simple models shown in Fig. 4. These are based on the amine product yielded from a reaction proceeding via an *E* or *Z* transition state and catalysed by the (*R*)-CPA. The opposite enantiomer will be formed if the (*S*)-CPA is employed as the catalyst. As a first case study, we evaluated fifteen additional reactions involving enecarbamates, a nucleophile not contained in the training set, and benzoyl imines, an imine subclass that is part of our initial training set[32] (Fig. 4). Each result was predicted using the comprehensive model, with an average absolute $\Delta\Delta G^{\ddagger}$ error of 0.37 kcal mol^{-1} (13 examples within 5% enantiomeric excess) and the absolute stereochemistry correctly assigned as *R*, demonstrating the ability of the model to extrapolate effectively to a new nucleophile. A slightly improved outcome is observed using the *E*-imine mechanistic model with an average error of 0.24 kcal mol^{-1} (all examples within 5% enantiomeric excess).

As the second case study, the hydrogenation of alkynyl ketimines catalysed by H8-BINOL where the 3,3′ groups = 3,5-$(CF_3)_2C_6H_3$ was predicted[33]. This is a more challenging scenario as both imine and catalyst components are not included in the training set. Again, accurate prediction of the outcomes was construed using the *Z*-imine mechanistic model, with an average absolute error of 0.30 kcal mol^{-1} and 13 examples predicted within 2% enantiomeric excess (Fig. 4). The stereochemical outcome was correctly determined to be *R* with the (*S*)-catalyst. Although the comprehensive model assesses the mechanistic scenario and therefore assigns the stereochemical outcome, it was not as accurate because the nucleophile information was categorical (symmetrical or displaced). Therefore, the beneficial effect of a large nucleophile for a *Z* reaction was not adequately captured.

These examples showcase that the predictive capabilities of the model are not limited to classifying the vast literature, but can be applied to analyse and predict new reactions even in situations where multiple components are varied.

As a final case study, we evaluated a recently reported reaction that was rendered highly predictable by application of machine learning algorithms. The study reported by Denmark and co-workers[34] involved the addition of thiols to benzoyl imines, a distinct reaction included in our training set. To utilize machine learning approaches, they performed 2,150 separate experiments using 43 catalysts to yield 25 different products (5 × 5 nucleophile/electrophile matrix). We postulated that our approach could reliably predict their results, including the best catalyst, TCYP (2,4,6-tricyclohexyl phenyl phosphoric acid), a CPA that is not in our training set. To test this hypothesis, all experimental results of this reaction type were removed from our original training data, the model was retrained, and deployed to predict their new dataset (34 reactions) collected with the best catalyst, TCYP. We conclude that our model—which lacks experimental data on this reaction—can also predict the enantioselectivities (average absolute $\Delta\Delta G^{\ddagger}$ error = 0.65 kcal mol^{-1} comprehensive model (26 examples within 5% enantiomeric excess), 0.67 kcal mol^{-1} *E*-imine-only model (25 examples within 5% enantiomeric excess)), confidently determining the stereochemical outcome to be *R* and TCYP to be a highly selective catalyst. Overall, through the combination of results generated from the out-of-sample prediction platforms, we can conclude that the *E*- and *Z*-focused correlations generate more accurate predictions but that the comprehensive model is valuable because it determines which equation should be deployed.

Here we have introduced a workflow with which to model enantioselectivity in assorted catalytic systems. The value of this approach is that complicated reaction conditions can be accounted for and successfully evaluated for multiple and diverse reactions. The ability to correlate and predict enantioselectivity using a single model that covers many reactions suggests that general transition-state features are fundamentally similar across the reaction range, allowing the transfer of observed reaction conditions from one reaction to another. This finding suggests a probable general phenomenon in asymmetric catalysis, whereby various transformations may be found to perform in the same manner when exposed to similar reaction conditions. Through the development of mechanism-specific correlations, such reaction similarities and reaction-specific mechanistic principles may be revealed.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1384-z.

Received: 24 February 2019; Accepted: 29 May 2019;
Published online 17 July 2019.

1. Houk, K. N. & Cheong, P. H.-Y. Computational prediction of small-molecule catalysts. *Nature* **455**, 309–313 (2008).
2. Davis, H. J. & Phipps, R. J. Harnessing non-covalent interactions to exert control over regioselectivity and site-selectivity in catalytic reactions. *Chem. Sci.* **8**, 864–877 (2017).
3. Knowles, R. R. & Jacobsen, E. N. Attractive noncovalent interactions in asymmetric catalysis: links between enzymes and small molecule catalysts. *Proc. Natl Acad. Sci. USA* **107**, 20678–20685 (2010).
4. Sigman, M. S., Harper, K. C., Bess, E. N. & Milo, A. The development of multidimensional analysis tools for asymmetric catalysis and beyond. *Acc. Chem. Res.* **49**, 1292–1301 (2016).
5. Ahneman, D. T., Estrada, J. G., Lin, S., Dreher, S. D. & Doyle, A. G. Predicting reaction performance in C-N cross-coupling using machine learning. *Science* **360**, 186–190 (2018).
6. Chuang, K. V. & Keiser, M. J. Comment on "Predicting reaction performance in C-N cross-coupling using machine learning". *Science* **362**, eaat8603 (2018).
7. Estrada, J. G., Ahneman, D. T., Sheridan, R. P., Dreher, S. D. & Doyle, A. G. Response to 'Comment on "Predicting reaction performance in C-N cross-coupling using machine learning"'. *Science* **362**, eaat8763 (2018).
8. Robbins, D. W. & Hartwig, J. F. A simple, multidimensional approach to high-throughput discovery of catalytic reactions. *Science* **333**, 1423–1427 (2011).
9. McNally, A., Prier, C. K. & MacMillan, D. W. C. Discovery of an alpha-C-H arylation reaction using the strategy of accelerated serendipity. *Science* **334**, 1114–1117 (2011).
10. Neel, A. J., Milo, A., Sigman, M. S. & Toste, F. D. Enantiodivergent fluorination of allylic alcohols: dataset design reveals structural interplay between achiral directing group and chiral anion. *J. Am. Chem. Soc.* **138**, 3863–3875 (2016).
11. Walsh, P. J. & Kozlowski, M. C. *Fundamentals of Asymmetric Catalysis* (University Science Books, 2008).
12. Yoon, T. P. & Jacobsen, E. N. Privileged chiral catalysts. *Science* **299**, 1691–1693 (2003).
13. Yamamoto, H. *Lewis Acids in Organic Synthesis* (Wiley, 2000).
14. Akiyama, T. Stronger Brønsted acids. *Chem. Rev.* **107**, 5744–5758 (2007).
15. Collins, K. D. & Glorius, F. Intermolecular reaction screening as a tool for reaction evaluation. *Acc. Chem. Res.* **48**, 619–627 (2015).
16. Gesmundo, N. J. et al. Nanoscale synthesis and affinity ranking. *Nature* **557**, 228–232 (2018).
17. Reetz, M. T. Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions. *Angew. Chem. Int. Ed.* **50**, 138–174 (2011).
18. Hansen, E., Rosales, A. R., Tutkowski, B., Norrby, P.-O. & Wiest, O. Prediction of stereochemistry using Q2MM. *Acc. Chem. Res.* **49**, 996–1005 (2016).
19. Metsänen, T. T. et al. Combining traditional 2D and modern physical organic-derived descriptors to predict enhanced enantioselectivity for the key aza-Michael conjugate addition in the synthesis of PrevymisTM (letermovir). *Chem. Sci.* **9**, 6922–6927 (2018).
20. Robak, M. T., Herbage, M. A. & Ellman, J. A. Synthesis and applications of tert-butanesulfinamide. *Chem. Rev.* **110**, 3600–3740 (2010).
21. Kobayashi, S., Mori, Y., Fossey, J. S. & Salter, M. M. Catalytic enantioselective formation of C–C bonds by addition to imines and hydrazones: a ten-year update. *Chem. Rev.* **111**, 2626–2704 (2011).
22. Nugent, T. C. *Chiral Amine Synthesis: Methods, Developments and Applications* (Wiley, 2010).
23. Silverio, D. L. et al. Simple organic molecules as catalysts for enantioselective synthesis of amines and alcohols. *Nature* **494**, 216–221 (2013).
24. Parmar, D., Sugiono, E., Raja, S. & Rueping, M. Complete field guide to asymmetric BINOL-phosphate derived Brønsted acid and metal catalysis: history and classification by mode of activation; Brønsted acidity, hydrogen bonding, ion pairing, and metal phosphates. *Chem. Rev.* **114**, 9047–9153 (2014).
25. Simón, L. & Goodman, J. M. Theoretical study of the mechanism of Hantzsch ester hydrogenation of imines catalyzed by chiral BINOL-phosphoric acids. *J. Am. Chem. Soc.* **130**, 8741–8747 (2008).
26. Reid, J. P., Simón, L. & Goodman, J. M. A practical guide for predicting the stereochemistry of bifunctional phosphoric acid catalyzed reactions of imines. *Acc. Chem. Res.* **49**, 1029 (2016).
27. Santiago, C. B., Guo, J.-Y. & Sigman, M. S. Predictive and mechanistic multivariate linear regression models for reaction development. *Chem. Sci.* **9**, 2398–2412 (2018).
28. Reid, J. P. & Sigman, M. S. Comparing quantitative prediction methods for the discovery of small-molecule chiral catalysts. *Nat. Rev. Chem.* **2**, 290–305 (2018).
29. Denmark, S. E., Gould, N. D. & Wolf, L. M. A systematic investigation of quaternary ammonium ions as asymmetric phase-transfer catalysts. Application of quantitative structure activity/selectivity relationships. *J. Org. Chem.* **76**, 4337–4357 (2011).
30. Hansch, C. & Leo, A. *Exploring QSAR: Fundamentals and Applications in Chemistry and Biology* (ACS, 1995).
31. Reid, J. P. & Goodman, J. M. Goldilocks catalysts: computational insights into the role of the 3,3′ substituents on the selectivity of BINOL-derived phosphoric acid catalysts. *J. Am. Chem. Soc.* **138**, 7910–7917 (2016).
32. Terada, M., Machioka, K. & Sorimachi, K. High substrate/catalyst organocatalysis by a chiral Brønsted acid for an enantioselective aza-ene-type reaction. *Angew. Chem. Int. Ed.* **45**, 2254–2257 (2006).
33. Chen, M.-W. et al. Organocatalytic asymmetric reduction of fluorinated alkynyl ketimines. *J. Org. Chem.* **83**, 8688–8694 (2018).
34. Zahrt, A. F. et al. Prediction of higher-selectivity catalysts by computer-driven workflow and machine learning. *Science* **363**, eaau5631 (2019).

METHODS

After the database of the reactions was constructed, the experimental output—enantiomeric ratios—were mathematically modelled through linear regression techniques to reveal which of the proposed parameters allow for the prediction of new outcomes. The detailed acquisition of parameters and the descriptor tables can be found in the Supplementary Information. The models produced were evaluated for their goodness of fit, R^2, and their robustness is demonstrated by external validation of the goodness of fit, the predicted R^2. The nearer the R^2 and slope values are to 1 (indicating a tight, one-to-one correlation between predicted and measured outcomes) and the nearer the intercept is to zero (indicating minimal systematic error), the more robust the model. Potential models were refined through number of parameters, because this allows for a mechanistically informative interrogation and cross-validation scores. LORO analysis was performed to probe general mechanistic principles, which provides the basis for mechanistic transfer of experimental observations and tested further by predicting out-of-sample.

Data availability

All data relating to this study is available in the Supplementary Information.

Code availability

All code used for model development is available in the Supplementary Information.

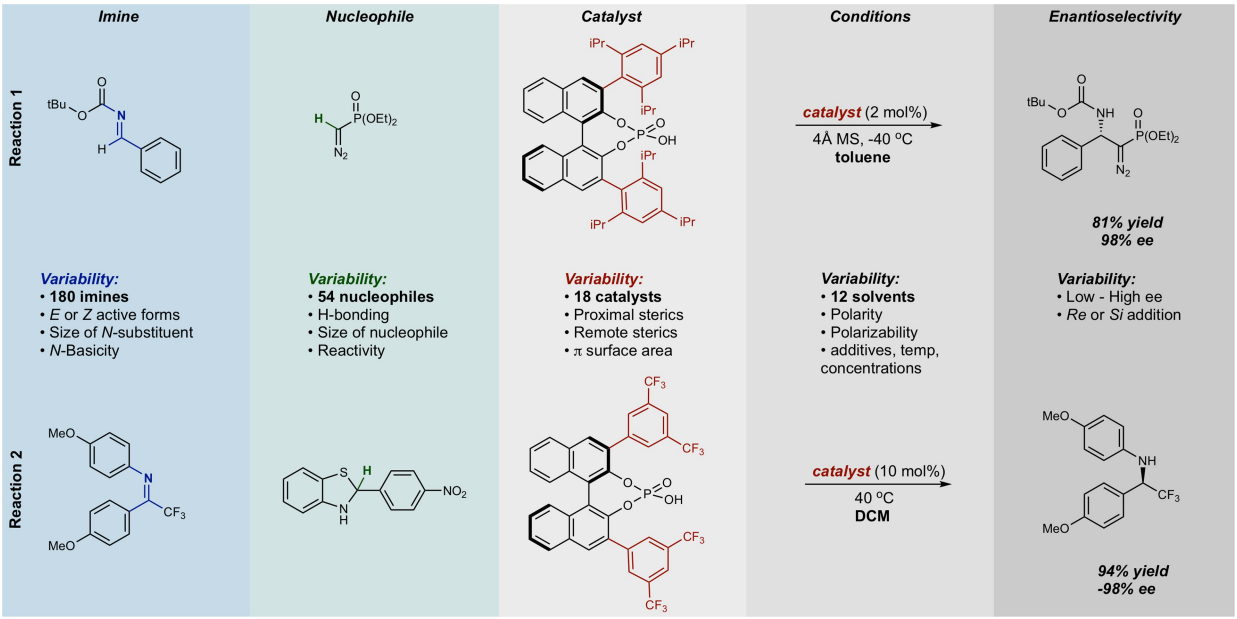
Acknowledgements J.P.R. thanks the EU Horizon 2020 Marie Skłodowska-Curie Fellowship (grant 792144) and M.S.S. thanks the NIH (grant GM-121383) for support of this work. Computational resources were provided from the Center for High Performance Computing (CHPC) at the University of Utah and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by the NSF (grant ACI-1548562) and provided through allocation TG-CHE180003.

Author contributions J.P.R. designed and performed all computations and statistical analyses. Both authors contributed to the analysis and writing of the manuscript.

Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1384-z.
Peer review information *Nature* thanks Timothy Cernak, Per-Ola Norrby and Robert Paton for their contribution to the peer review of this work.
Correspondence and requests for materials should be addressed to M.S.S.



Extended Data Fig. 1 | Reaction component comparison. Parameterization challenges for the identification of numerical descriptors in reaction dimension, demonstrated using two reactions that represent the extremes of multidimensional feature space. MS, molecular sieves.

ARTICLE

https://doi.org/10.1038/s41586-019-1385-y

Comprehensive single-cell transcriptome lineages of a proto-vertebrate

Chen Cao[1,5], Laurence A. Lemaire[1,5], Wei Wang[2], Peter H. Yoon[3], Yoolim A. Choi[3], Lance R. Parsons[1], John C. Matese[1], Wei Wang[1], Michael Levine[1,3]* & Kai Chen[1,4]*

Ascidian embryos highlight the importance of cell lineages in animal development. As simple proto-vertebrates, they also provide insights into the evolutionary origins of cell types such as cranial placodes and neural crest cells. Here we have determined single-cell transcriptomes for more than 90,000 cells that span the entirety of development—from the onset of gastrulation to swimming tadpoles—in *Ciona intestinalis*. Owing to the small numbers of cells in ascidian embryos, this represents an average of over 12-fold coverage for every cell at every stage of development. We used single-cell transcriptome trajectories to construct virtual cell-lineage maps and provisional gene networks for 41 neural subtypes that comprise the larval nervous system. We summarize several applications of these datasets, including annotating the synaptome of swimming tadpoles and tracing the evolutionary origin of cell types such as the vertebrate telencephalon.

Single-cell RNA-sequencing methods are revolutionizing our understanding of how cells are specified to become definitive tissues during development[1–5]. These studies allow the elucidation of virtual lineages for select tissues, and provide detailed expression profiles for cell types such as pluripotent progenitor cells. However, a limitation of previous studies has been the incomplete coverage of vertebrate embryos, owing to the large numbers of cells present in these embryos.

As one of the closest living relatives of vertebrates[6], the ascidian *C. intestinalis* serves a critical role in understanding developmental and physiological processes that are comparable to—but far less complex than—those of vertebrates. In comparison to vertebrate embryos, ascidian embryos are simple: gastrulating embryos are composed of only 100–200 cells, and swimming tadpoles contain about 2,500 cells. Owing to these small numbers of cells, it is possible to obtain comprehensive coverage of every cell type during development, including rare neuronal subtypes.

Here we extend insights into the regulatory 'blueprint' that spans the early phases of embryogenesis[7] by profiling the transcriptomes of individual cells in sequentially staged *Ciona* embryos, from gastrulation at the 110-cell stage to the neurula and larval stages. Reconstructed temporal expression profiles reveal the specification and differentiation of individual cell types. Nearly 40 subtypes of neurons were identified, even though the central nervous system of the *Ciona* larva is composed of only 177 neurons[8]. The resulting high-resolution transcriptome trajectories, regulatory cascades and provisional gene networks provide insights into the evolution of novel cell types in vertebrates, including those of the telencephalon.

Specification of cell fate

Synchronized embryos from ten different stages of development were rapidly dissociated in RNase-free calcium-free synthetic seawater, and individual cells were processed in the 10x Genomics Chromium system with at least two biological replicates for each developmental stage (Fig. 1a, Extended Data Fig. 1, Supplementary Table 1, Methods). The staged embryos span all of the hallmark processes of development, beginning with gastrulation and culminating in swimming tadpoles (at which point all larval cell types, tissues and organs are formed) (Fig. 1b). In total, we profiled 90,579 cells, which corresponds to an average of over 12-fold coverage for every cell across each of the sampled stages (Supplementary Table 1). Individual cells were sequenced to an average depth of about 12,000 unique molecular identifiers, which enabled the recovery of rare populations such as germ cells (which constitute about 0.1% of cells in swimming tadpoles).

t-distributed stochastic neighbour-embedding (*t*-SNE) projections of the transcriptomes at all ten stages of development identified coherent clusters of individual tissues, including heart, tail muscles, endoderm, notochord, germ cells, epidermis, nervous system and mesenchyme (Extended Data Fig. 2a–l). Several tissues—such as the nervous system and mesenchyme—exhibit a progressive increase in cell complexity during development (Extended Data Fig. 2c–l), which results in the appearance of more cell clusters at later stages of embryogenesis. We also found that most individual tissues displayed less variation in their transcriptome profiles during development, when compared with divergent cell types at the same time points. This is particularly evident for the developing germ line, because it is transcriptionally quiescent during the time frame of our analysis[9].

The specific and stable expression of cell-specific marker genes (Extended Data Fig. 2m, Supplementary Table 2), such as *Brachyury* for the developing notochord and *Twist-like-2* for the mesenchyme[10,11], facilitated the reconstruction of temporal profiles for different tissues. This study also identified a variety of genes, including *Kdm8* (a histone H3K36me2 demethylase expressed in mesenchyme lineages), as tissue-specific markers.

Classical cell-lineage studies suggest that all of the major tissues of the ascidian tadpole are specified before gastrulation, at the 110-cell stage[12] (Fig. 1c). Most of the internal organs—including the notochord, endoderm, tail muscles, heart, germ cells, and regions of the nervous system—are derived from vegetal lineages. By contrast, animal blastomeres give rise to ectodermal derivatives, including epidermis and associated sensory neurons, and regions of the nervous system. Each of these cell types was identified as a discrete cluster in the *t*-SNE projections of dissociated 110-cell embryos (Fig. 1d).

Several tissues are already seen to segregate into distinct anterior and posterior clusters by the 110-cell stage, including the notochord,

[1]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [2]Stowers Institute for Medical Research, Kansas City, MO, USA. [3]Department of Molecular Biology, Princeton University, Princeton, NJ, USA. [4]Present address: The Yunnan Key Laboratory of Primate Biomedical Research, Institute of Primate Translational Medicine, Kunming University of Science and Technology, Kunming, China. [5]These authors contributed equally: Chen Cao, Laurence A. Lemaire. *e-mail: msl2@princeton.edu; chenk@lpbr.cn

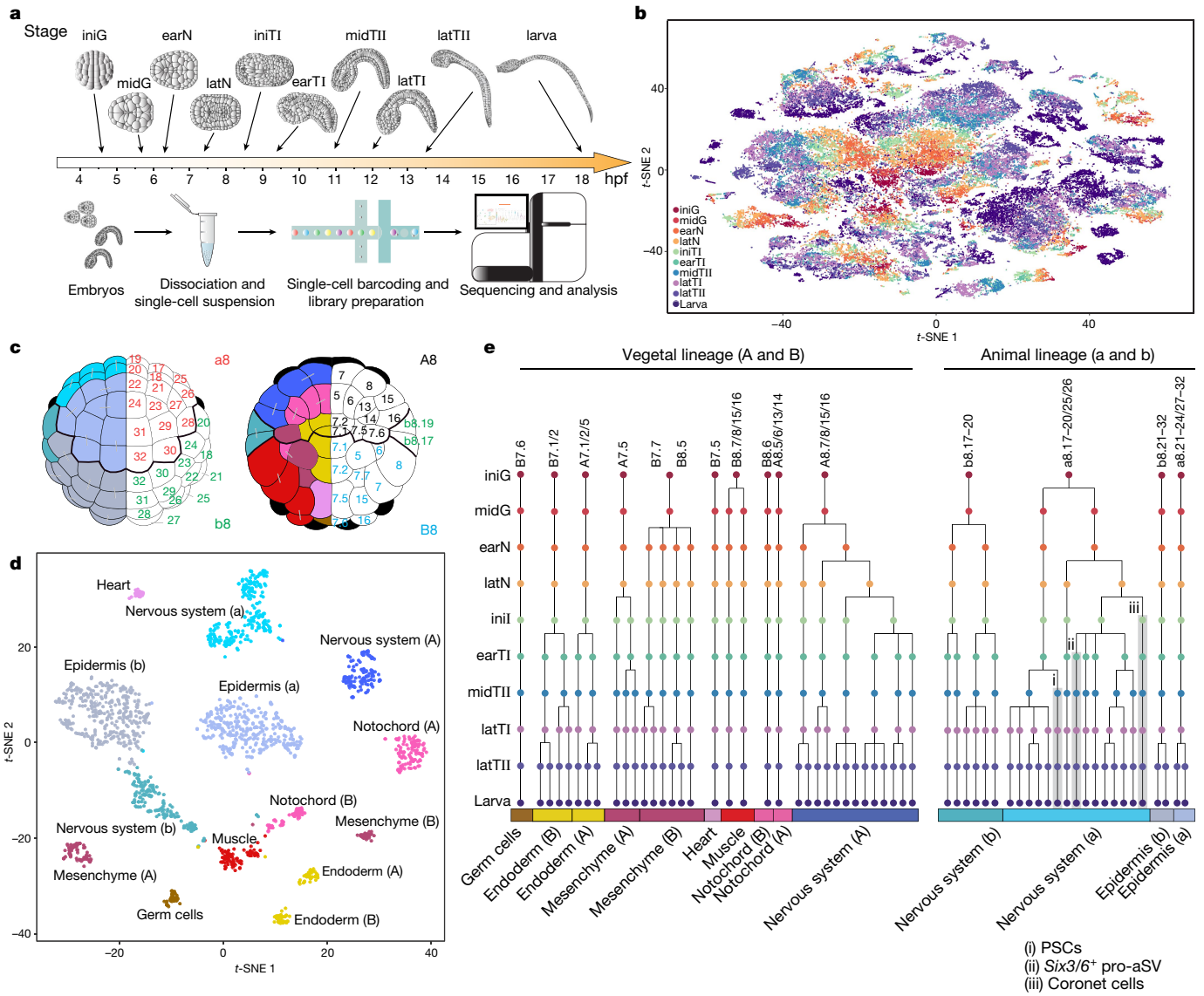


Fig. 1 | Overview of single-cell RNA-sequencing assays and cell-type specification at the onset of gastrulation. a, Staged embryos were collected from ten developmental stages: the initial gastrula (iniG), mid-gastrula (midG), early neurula (earN), late neurula (latN), initial tailbud I (iniTI), early tailbud I (earTI), mid-tailbud II (midTII), late tailbud I (latTI), late tailbud II (latTII) and larval (larva) stages. $n = 2$ biological replicates per stage (iniG to latTII stages); $n = 3$ biological replicates (larval stage). hpf, hours post-fertilization. **b**, *t*-SNE plot of the entire dataset ($n = 90{,}579$ cells). Cells are colour-coded according to developmental stage. **c**, Schematics of animal (left) and vegetal (right) blastomeres of a *Ciona* embryo at the initial gastrula stage. Tissue types are colour-coded (left) and named according to Conklin's nomenclature (right). Bold lines indicate the boundaries between the blastomeres of lineages a, b, A and B. **d**, *t*-SNE plot of transcriptomes from single cells at the initial gastrula stage ($n = 1{,}731$ cells) using the colour-coding scheme shown in **c**. Each of the major tissues maps within a separate cluster. **e**, Virtual lineage trees were reconstructed using transcriptome profiles from sequential developmental stages. The points in the tree represent inferred developmental transitions from initial gastrula to larva. Only unambiguous alignments are shown (Methods). Branches with shadows represent lineages of palp sensory cells (PSCs) (i), *Six3/6*$^{+}$ pro-anterior sensory vesicle (pro-aSV) (ii) and coronet cells (iii).

endoderm and lateral plate sensory cells. These observations validate and extend classical evidence for the specification of all major larval tissues at the 110-cell stage. Our expression profiles of individual cell types revealed previously known and newly identified potential fate determinants (Extended Data Fig. 3, Supplementary Table 2): for example, *Irx-B* is specifically expressed in a-lineage (anterior) epidermis and *Not* is expressed in the b-lineage (posterior).

Reconstructing cell lineages

The alignment of transcriptome profiles of individual cell types at sequential stages of development enabled the reconstruction of virtual lineage maps (Fig. 1e, Methods). In total, 60 cell types were identified in swimming tadpoles, and the corresponding virtual lineages of these cell types could be traced to blastomeres at the 110-cell stage (the time of fate restriction). The reconstructed lineages are in close agreement with known lineage information, and provide insights into the specification and differentiation of individual cell types. For example, the transcriptome profiles accurately capture the muscle and heart lineages (Extended Data Fig. 4a, b), as well as the primary (from A8 blastomeres) and secondary (from B8 blastomeres) lineages of the notochord[12] (Extended Data Fig. 5). The mesenchyme has previously been shown to be derived from three separate lineages (from A7.6, B7.7 and B8.5 blastomeres)[11], and our analyses suggest they segregate to produce nine cell types (Extended Data Fig. 4c, d). Similarly, the head and trunk endoderm produce seven cell types (Extended Data Fig. 4e). This is a considerably higher level of resolution than that obtained by conventional experimental studies[12,13].

The transcriptome maps also capture more-nuanced lineage information. For example, dopaminergic neurons (coronet cells) of the central nervous system were found to share a common lineage with the

pro-anterior sensory vesicle, the anterior-most terminus of the neural tube that fuses with the stomodeum to form the neuropore[14]. Both derivatives share a common origin with palp sensory cells, which arise from the non-neural proto-placodal territory located immediately anterior of the neural tube (Fig. 1e)—this is consistent with the model for the evolution of the vertebrate telencephalon discussed in 'Evolution of cell types' below.

Transitional properties of the notochord

The notochord is a derivative of the mesoderm, and is a defining innovation of chordates[15]. However, the notochord exhibits distinctive properties in cephalochordates and vertebrates. Cephalochordates such as *Amphioxus* contain a muscular notochord that helps to power movements of the tail[16], whereas the vertebrate notochord is non-muscular and provides structural support for derivatives of the paraxial mesoderm. The *Ciona* notochord appears to contain a mixture of both properties.

The primary (A-lineage) and secondary (B-lineage) notochord cells are clearly resolved into subclusters throughout development (Extended Data Fig. 5a). By constructing single-cell trajectories, it was possible to identify cell signalling and regulatory genes in each lineage (Extended Data Fig. 5b, c). In addition to the identification of genes that are known to be differentially expressed in the two lineages (such as *ZicL* and *Notch*)[17,18], we were able to identify distinctive regulatory strategies for the two lineages (Extended Data Fig. 5b, c). For example, *Otx* and *Not* are specifically expressed in the secondary notochord, along with the muscle determinants *Tbx6a*, *Tbx6c* and *Tbx6d*[19] (Extended Data Fig. 3). They precede expression of muscle identity genes such as calsequestrin (*Casq1/2*; solidi in *Ciona* gene symbols separate multiple vertebrate homologues (as *Ciona* has not undergone genome duplication)), myosin (*Mlra/Mlrv/Myl5*) and tropomyosin (*Tpm1/2/3*) (Extended Data Fig. 5d, Supplementary Table 2). None of these genes is expressed in the primary notochord[20]. Moreover, the 5′ regulatory regions of these genes contain clusters of Tbx6 binding motifs (Supplementary Table 3), which suggests their direct regulation by muscle determinants. Gene reporter assays verified restricted expression of *Casq1/2* and *KH.C9.405* (Supplementary Table 2) in the secondary notochord and tail muscles (Extended Data Fig. 5e). It therefore appears that a muscle differentiation program is purposefully deployed in the secondary, but not the primary, notochord. These developmental programs suggest that *Ciona* possesses properties of both the notochords seen in cephalochordates and those of vertebrates.

Identification of individual neurons

The central nervous system of swimming tadpoles is composed of only 177 neurons[8], which allows for the reconstruction of detailed transcriptome trajectories for individual neurons (Methods). We profiled 22,198 neural cells derived from the a-, b- and A-lineages (Extended Data Fig. 6a–c) across all 10 stages of development. This represents an average of about sevenfold coverage for every cell type (Supplementary Table 4). A total of 41 neural derivatives were identified in swimming tadpoles (Fig. 2, Extended Data Fig. 6d). These cells map to different regions of the nervous system, including the sensory vesicle, motor ganglion, nerve cord, peripheral sensory cells and associated interneurons. Distinctive combinations of regulatory genes were identified in the neural subtypes (Fig. 2, Extended Data Fig. 6e, Supplementary Table 2). For example, coronet cells are the only dopaminergic neurons in the *Ciona* central nervous system. Coronet cells express high levels of *Ptf1a* and *Meis*, which are sufficient to reprogram the central nervous system into supernumerary coronet cells[21]. It is possible that other combinations of cell-specific transcription factors specify additional neural subtypes (for example, *Bsh*, *Lhx2/9* and *Aristaless* in the anterior sensory vesicle).

The high coverage of individual transcriptomes enabled the identification of rare neuronal subtypes (Extended Data Fig. 7). For example, there are only two pairs of bipolar tail neurons in swimming tadpoles[22], and these were found to express galanin and two of its receptors (*Galr1* and *Galr2*) (Supplementary Table 2). Galanin has previously been

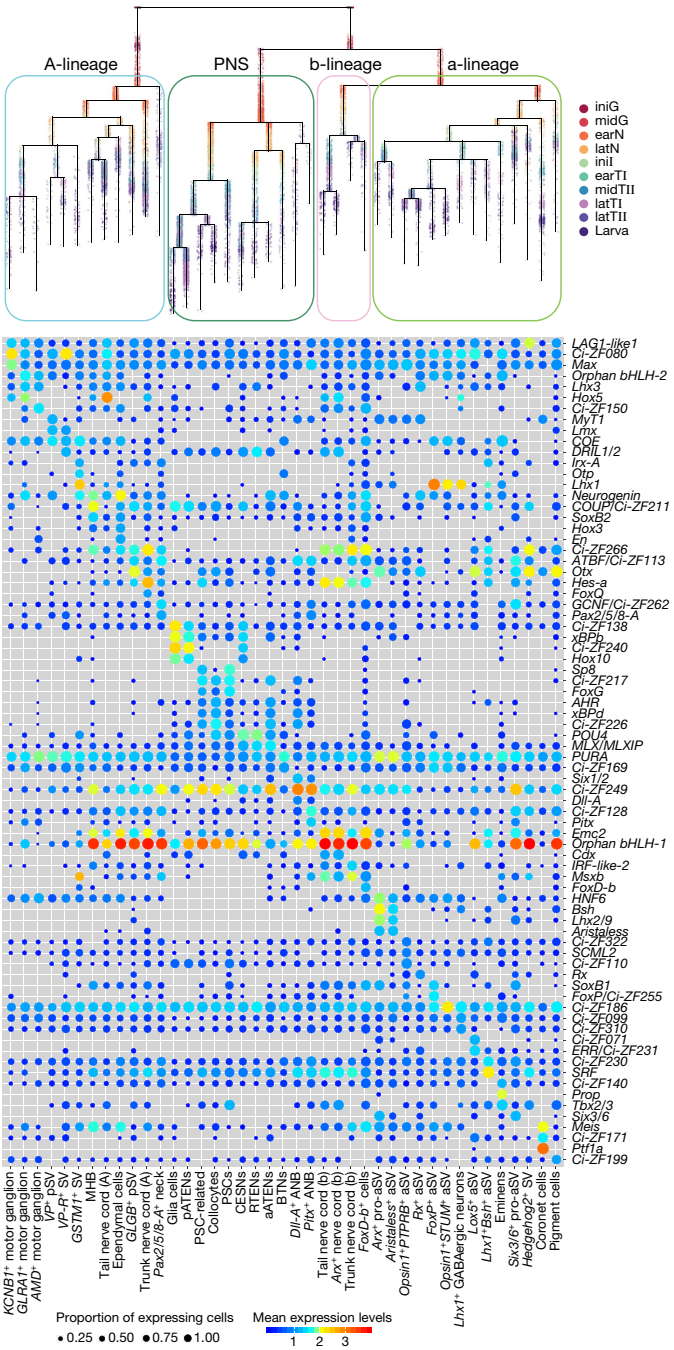


Fig. 2 | Transcriptome trajectories for defined individual neurons. Reconstructed expression lineage for the entire nervous system. Top, cells are coloured by developmental stage, and the a-lineage, b-lineage and A-lineage branches of the central nervous system and peripheral nervous system (PNS) are identified. Cells are ordered by pseudotime along each trajectory. Bottom, dot plot of the top three most-highly expressed regulatory genes in each neural subcluster at the larval stage. Dot size represents the percentage of cells that express the transcription factor, and the dot colour shows the averaged level of expression. aATENs, anterior apical trunk epidermal neurons; ANB, anterior neural boundary; aSV, anterior sensory vesicle; BTNs, bipolar tail neurons; CESNs, caudal epidermal sensory neurons; MHB, midbrain–hindbrain boundary; pATENs, posterior apical trunk epidermal neurons; RTENs, rostral trunk epidermal neurons; SV, sensory vesicle. Letters in parentheses denote lineages.

implicated in neuro-regeneration and axogenesis[23,24]. A reporter gene that contains *Galr2* regulatory sequences mediates restricted expression in the bipolar tail neurons (Extended Data Fig. 7a). Similarly, a pair of decussating neurons—which have a central role in the startle response

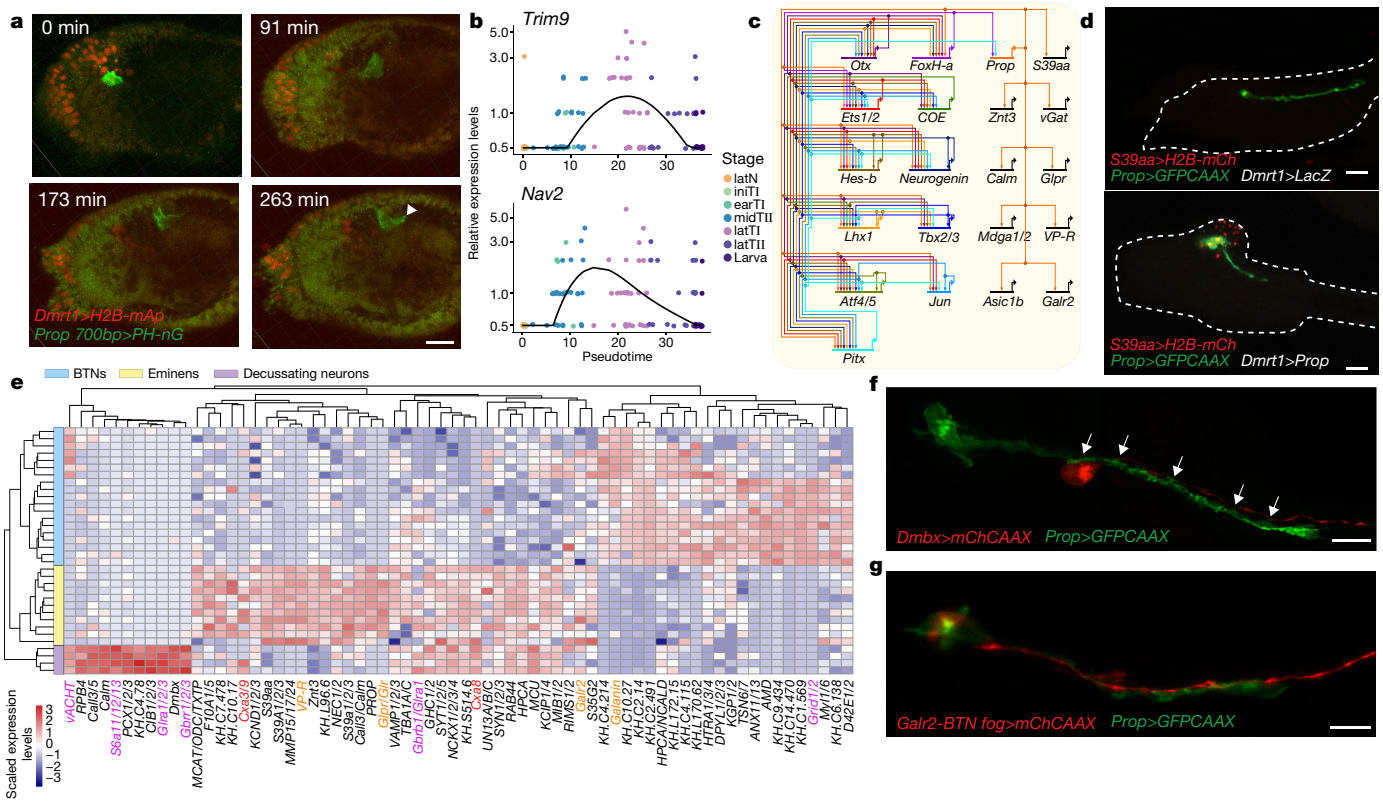


Fig. 3 | Integration of transcriptome maps and synaptome neuronal circuits. a, Snap shots of a time-lapse video of Eminens cell migration (from latTI to latTIII, $n = 2$ embryos). The embryo expressed H2B–mApple (H2B–mAp) under the regulatory sequences of *Dmrt1* (red), and PH–mNeonGreen (PH-nG) driven by *Prop* regulatory sequences (green). The Eminens cell migrates from the anterior side towards the posterior of the sensory vesicle before axogenesis (arrowhead). Numbers in top left indicate elapsed time in minutes. **b**, Pseudo-temporal expression of two migration-related genes from the early neurula to the larval stage. **c**, A provisional gene regulatory network of Eminens cells is shown, based on the regulatory cascade (Extended Data Fig. 9a). Only representative regulatory genes are shown (Methods). **d**, Upon overexpression of *Prop* in the anterior neural plate, supernumerary *S39aa*$^{+}$ cells (H2B–mCherry (H2B–mCh), red) are observed (bottom, *Dmrt1>Prop*) compared with control embryos (top, *Dmrt1>LacZ*). The embryos also expressed a membrane GFP as a *Prop* reporter gene (GFPCAAX, green, $n = 3$ electroporation experiments for both conditions). **e**, Heat map of differentially expressed genes between Eminens neurons, decussating neurons and bipolar tail neurons. Genes involved in neurotransmission are coloured in magenta, connexin genes in red and neuropeptide-associated genes in orange. **f**, Em2 has multiple contacts (arrows) with the decussating neurons, as observed by double labelling with *Prop* and *Dmbx* reporter genes (green and red, respectively). $n = 3$ electroporation experiments. **g**, Double labelling with *Prop* and bipolar-tail-neuron-specific *Galr2* reporter genes (green and red, respectively) shows extensive contacts of bipolar tail neuron axon with Eminens neurons. $n = 2$ electroporation experiments. Scale bars, 20 μm (**a**, **d**), 10 μm (**f**, **g**).

of vertebrates[25,26]—was unambiguously identified on the basis of their selective expression of the homeobox gene *Dmbx*[27] (Extended Data Fig. 7b).

Additional specific neuronal subtypes were identified on the basis of restricted expression of select marker genes. Posterior sensory vesicle neurons that are positive for the vasopressin/oxytocin (*VP*) gene express several neuropeptides (Supplementary Table 5), including *VP* and an uncharacterized neuropeptide[28] (*NP*; see Supplementary Table 6). *NP* expression is restricted to a small group of neurons located in the posterior-most regions of the sensory vesicle (Extended Data Fig. 7c, e–f). A pair of Eminens neurons was identified by expression of a reporter gene containing *Prop* 5′ regulatory sequences (Extended Data Fig. 7d). These studies document the feasibility of identifying the transcriptome trajectories and virtual lineages of individual defined neurons, including rare subtypes such as Eminens neurons.

Transcriptome and synaptome integration

The recently reported *Ciona* synaptome identified a single Eminens neuron (Em2) as a key regulator of decussating neurons[8,25]. The pair of Eminens neurons was identified in our datasets on the basis of their expression of marker genes of GABAergic (γ-aminobutyric-acid-releasing) neurons and *Prop* (Extended Data Fig. 8a, b). Moreover, reporter genes that contain *Prop* regulatory sequences are selectively expressed in a pair of neurons that display all of the properties of Eminens neurons, including morphology and location[25,29–31] (Extended Data Fig. 7d). Transcriptome trajectories of Eminens neurons suggest that they arise from the a-lineage (Fig. 2), even though they are located in the posterior regions of the sensory vesicle. This apparent discrepancy was resolved by live-cell imaging. We found that Eminens neurons undergo long-range migration from the forebrain to posterior regions of the sensory vesicle (Fig. 3a, Supplementary Video 1). These movements correlate with the expression of a variety of genes that are implicated in migration and axogenesis, including *Nav2* and *Trim9*[32,33] (Fig. 3b).

Regulatory cascades of cell-signalling components and transcription factors enabled the formulation of a provisional gene regulatory network for the Eminens neurons (Fig. 3c, Extended Data Fig. 9a). The lynchpin of this network is *Prop*, a homeobox gene that appears to regulate a variety of genes that are involved in neuronal function; these include neuropeptide receptors (*VP-R*, *Glpr* and *Galr2*), zinc neuromodulation (*Znt3* and *S39aa*) and GABAergic markers (*vGat*) (Fig. 3c, Extended Data Fig. 8). Support for this network was obtained by manipulating a minimal *Prop* enhancer. Point mutations in the binding site for *FoxH-a*, one of the predicted upstream regulators of *Prop*, caused a significant (Fisher's exact test, $P = 1.27 \times 10^{-7}$) reduction in the expression of the minimal *Prop* reporter gene (Extended Data Fig. 9b, c). More importantly, overexpression of *Prop* in anterior regions of the sensory vesicle (via a *Dmrt1>Prop* fusion gene) resulted in the

formation of supernumerary Eminens neurons and ectopic activation of downstream reporter genes (for example, *S39aa*) (Fig. 3d, Extended Data Fig. 9d, e).

We next sought to leverage this information to gain insights into the neuronal interactions that underlie the startle response (Fig. 3e). A centrepiece of the startle circuit is the pair of decussating neurons, which correspond to the Mauthner neurons in the brain stem of fish and frogs[26]. The decussating neurons integrate a variety of sensory information to trigger a fast escape reflex. As predicted by previous studies[25,31], interactions between Em2 and the decussating neurons (Fig. 3f) are probably inhibitory, as Eminens neurons express GABAergic markers such as *vGat* and *Gad* (Extended Data Fig. 8a, b) whereas the decussating neurons express GABA receptors (Fig. 3e, Supplementary Table 2). The decussating neurons also express glutamate receptors (Supplementary Table 2), which suggests that they respond to tonic glutamate signals.

The transcriptome datasets further raise the possibility that the startle circuit may be modulated by secreted neuropeptides (Supplementary Table 5). Both Eminens and decussating neurons express receptors for galanin, which is expressed in the bipolar tail neurons (Fig. 3e, g, Extended Data Fig. 8h). The bipolar tail neurons have previously been likened to the dorsal root ganglia derivatives of the neural crest in vertebrates[22]. Galanin promotes survival of dorsal root ganglia neurons during development and after injury. It is possible that galanin serves as a tropic factor for Em2, because the bipolar tail neurons directly interact with the cell body of this neuron (Fig. 3g). Moreover, modulation of Em2 by additional neuropeptides is suggested by the fact that Em2 expresses a VP receptor. As shown above, VP^{+} cells express genes for a number of secreted neuropeptides—including *VP* and *NP* (Extended Data Fig. 7e). The VP^{+} cells are in close proximity with Em2 (Extended Data Fig. 7f).

Our transcriptome datasets provide substantive annotations of the neuronal circuits that have been described by recent synaptome studies[8,29], suggest both targeted growth and feedback inhibition of the startle response by bipolar tail neurons, and implicate neuropeptides (such as galanin and vasopressin/oxytocin) as potential modulators of the circuit, in addition to canonical neurotransmitters.

Evolution of cell types

Previous studies suggest that *Ciona* possesses the rudiments of key vertebrate innovations, including the neural crest, cranial placodes and the cardio-pharyngeal mesoderm[22,34–37]. However, the evolutionary origin of the telencephalon, which arises from the anterior-most regions of the forebrain, remains uncertain. The telencephalon contains the olfactory bulb and regions that control higher-order brain functions, such as the neocortex of humans. Forebrain regions of the *Ciona* central nervous system give rise to dopaminergic coronet cells and neuropore, but lack telencephalon derivatives such as the olfactory bulb.

To explore the origins of the telencephalon, we examined the gene-regulatory cascades for derivatives of the anterior-most regions of the neural plate, particularly palp sensory cells and the pro-anterior sensory vesicle (Extended Data Figs. 10–12, Methods). The palp sensory cells, also known as axial columnar cells[38], express a cascade of cell-signalling components and regulatory genes, including *FoxC*, *Dlx*, *FoxG*, *Isl* and *SP8* (Extended Data Fig. 12a, c, Supplementary Table 2). A similar regulatory cascade has previously been implicated in the specification of the telencephalon in vertebrates[39,40].

We also determined transcriptome trajectories for the pro-anterior sensory vesicle (the anterior-most regions of the neural tube), located adjacent to the proto-placodal territory that forms the palp sensory cells. The pro-anterior sensory vesicle first expresses anterior determinants (for example, *Otx*), followed by cell-specification genes such as *FoxJ1*, *Six1/2*, *Six3/6*, *Lhx2/9*, *Pitx* and *Otp* (Extended Data Fig. 12b, d, Supplementary Table 2). Many of these genes have also previously been implicated in the development of forebrain derivatives, including regions of the telencephalon[41,42].

We propose that the vertebrate telencephalon arose through the incorporation of non-neural ectoderm in anterior regions of the neural tube (Fig. 4a). To test this model, we examined the expression of a *Ciona FoxG* reporter gene in *Ciona* larvae and transgenic killifish (*Nothobranchius furzeri*) embryos (Fig. 4b, c, Methods). This reporter is expressed in palp cells of *Ciona* embryos (Fig. 4b). It also mediates expression in subsets of cells in the olfactory bulb of the killifish telencephalon (Fig. 4c), as well as in placodal derivatives such as the lens of the eye (Extended Data Fig. 12e). These observations are consistent with the incorporation of proto-placodal gene-regulatory modules (for example, axial columnar cells) into an expanded forebrain of vertebrates.

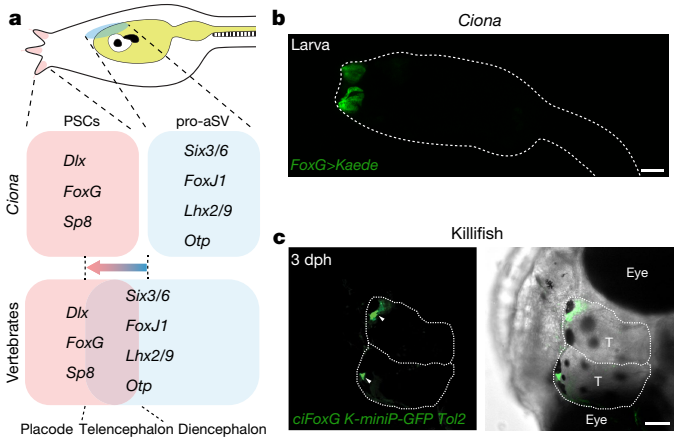


Fig. 4 | Model for the evolution of the telencephalon. a, Proposed model of the evolution of the vertebrate telencephalon. The telencephalon arose from the incorporation of anterior placodal gene-regulatory module into forebrain regions of the neural tube. Key regulatory components in *Ciona* palp sensory cells (including *Dlx*, *FoxG* and *Sp8*) and in the pro-anterior sensory vesicle (including *Six3/6*, *FoxJ1*, *Lhx2/9* and *Otp*) are conserved in the vertebrate telencephalon. **b**, The *FoxG* reporter gene (Kaede, green) exhibits restricted expression in palp sensory cells but not anterior regions of the sensory vesicle of *Ciona* larvae ($n = 2$ electroporation experiments). **c**, In killifish, GFP driven by the *Ciona FoxG* regulatory sequences and a zebrafish minimal promoter is expressed in a subset of cells in the olfactory bulb of the telencephalon (arrowheads, left). $n = 3$ independent transgenic lines (Methods). T, telencephalon. dph, days post-hatching. Scale bars, 20 μm (**b**), 250 μm (**c**).

In summary, we have presented comprehensive transcriptome trajectories, regulatory cascades and provisional gene networks for over 60 cell types (including nearly 40 neuronal subtypes) that comprise the *Ciona* tadpole. These datasets substantially extend classical lineage maps and regulatory blueprints, and provide a source of information for reconstructing the contributions of individual cells, lineages and tissues to critical morphogenetic processes, such as gastrulation, neurulation, notochord intercalation, tail elongation, compartmentalization of the gut and nervous system, and the formation of complex neuronal circuits that control behaviour. Our datasets also provide insights into the evolutionary transition between invertebrates and vertebrates, including the dual properties of the *Ciona* notochord and the expansion of the vertebrate forebrain. Current single-cell studies encompasses a broad spectrum of cell types and systems[1–5,43], offering unprecedented opportunities to trace the evolutionary origins of every cell, tissue and organ in the human body.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests are available at https://doi.org/10.1038/s41586-019-1385-y.

Received: 27 August 2018; Accepted: 10 June 2019;
Published online 10 July 2019.

1. Briggs, J. A. et al. The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018).
2. Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).

3. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
4. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
5. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496–502 (2019).
6. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006).
7. Imai, K. S., Levine, M., Satoh, N. & Satou, Y. Regulatory blueprint for a chordate embryo. *Science* **312**, 1183–1187 (2006).
8. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The CNS connectome of a tadpole larva of *Ciona intestinalis* (L.) highlights sidedness in the brain of a chordate sibling. *eLife* **5**, e16962 (2016).
9. Prodon, F., Yamada, L., Shirae-Kurabayashi, M., Nakamura, Y. & Sasakura, Y. Postplasmic/PEM RNAs: a class of localized maternal mRNAs with multiple roles in cell polarity and development in ascidian embryos. *Dev. Dyn.* **236**, 1698–1715 (2007).
10. Corbo, J. C., Levine, M. & Zeller, R. W. Characterization of a notochord-specific enhancer from the Brachyury promoter region of the ascidian, *Ciona intestinalis*. *Development* **124**, 589–602 (1997).
11. Tokuoka, M., Imai, K. S., Satou, Y. & Satoh, N. Three distinct lineages of mesenchymal cells in *Ciona intestinalis* embryos demonstrated by specific gene expression. *Dev. Biol.* **274**, 211–224 (2004).
12. Nishida, H. Cell lineage analysis in ascidian embryos by intracellular injection of a tracer enzyme. III. Up to the tissue restricted stage. *Dev. Biol.* **121**, 526–541 (1987).
13. Nakazawa, K. et al. Formation of the digestive tract in *Ciona intestinalis* includes two distinct morphogenic processes between its anterior and posterior parts. *Dev. Dyn.* **242**, 1172–1183 (2013).
14. Veeman, M. T., Newman-Smith, E., El-Nachef, D. & Smith, W. C. The ascidian mouth opening is derived from the anterior neuropore: reassessing the mouth/neural tube relationship in chordate evolution. *Dev. Biol.* **344**, 138–149 (2010).
15. Stemple, D. L. Structure and function of the notochord: an essential organ for chordate development. *Development* **132**, 2503–2512 (2005).
16. Suzuki, M. M. & Satoh, N. Genes expressed in the amphioxus notochord revealed by EST analysis. *Dev. Biol.* **224**, 168–177 (2000).
17. Yagi, K., Satou, Y. & Satoh, N. A zinc finger transcription factor, ZicL, is a direct activator of *Brachyury* in the notochord specification of *Ciona intestinalis*. *Development* **131**, 1279–1288 (2004).
18. Hudson, C. & Yasuo, H. A signalling relay involving Nodal and Delta ligands acts during secondary notochord induction in *Ciona* embryos. *Development* **133**, 2855–2864 (2006).
19. Yagi, K., Takatori, N., Satou, Y. & Satoh, N. *Ci-Tbx6b* and *Ci-Tbx6c* are key mediators of the maternal effect gene *Ci-macho1* in muscle cell differentiation in *Ciona intestinalis* embryos. *Dev. Biol.* **282**, 535–549 (2005).
20. Takahashi, H. et al. *Brachyury* downstream notochord differentiation in the ascidian embryo. *Genes Dev.* **13**, 1519–1523 (1999).
21. Horie, T. et al. Regulatory cocktail for dopaminergic neurons in a protovertebrate identified by whole-embryo single-cell transcriptomics. *Genes Dev.* **32**, 1297–1302 (2018).
22. Stolfi, A., Ryan, K., Meinertzhagen, I. A. & Christiaen, L. Migratory neuronal progenitors arise from the neural plate borders in tunicates. *Nature* **527**, 371–374 (2015).
23. Shi, T. J. et al. Sensory neuronal phenotype in galanin receptor 2 knockout mice: focus on dorsal root ganglion neurone development and pain behaviour. *Eur. J. Neurosci.* **23**, 627–636 (2006).
24. Holmes, F. E. et al. Targeted disruption of the galanin gene reduces the number of sensory neurons and their regenerative capacity. *Proc. Natl Acad. Sci. USA* **97**, 11563–11568 (2000).
25. Ryan, K., Lu, Z. & Meinertzhagen, I. A. Circuit homology between decussating pathways in the *Ciona* larval CNS and the vertebrate startle-response pathway. *Curr. Biol.* **27**, 721–728 (2017).
26. Korn, H. & Faber, D. S. The Mauthner cell half a century later: a neurobiological model for decision-making? *Neuron* **47**, 13–28 (2005).
27. Stolfi, A. & Levine, M. Neuronal subtype specification in the spinal cord of a protovertebrate. *Development* **138**, 995–1004 (2011).
28. Hamada, M. et al. Expression of neuropeptide- and hormone-encoding genes in the *Ciona intestinalis* larval brain. *Dev. Biol.* **352**, 202–214 (2011).
29. Ryan, K., Lu, Z. & Meinertzhagen, I. A. The peripheral nervous system of the ascidian tadpole larva: types of neurons and their synaptic networks. *J. Comp. Neurol.* **526**, 583–608 (2018).
30. Imai, J. H. & Meinertzhagen, I. A. Neurons of the ascidian larval nervous system in *Ciona intestinalis*: I. Central nervous system. *J. Comp. Neurol.* **501**, 316–334 (2007).
31. Takamura, K., Minamida, N. & Okabe, S. Neural map of the larval central nervous system in the ascidian *Ciona intestinalis*. *Zool. Sci.* **27**, 191–203 (2010).
32. Hekimi, S. & Kershaw, D. Axonal guidance defects in a *Caenorhabditis elegans* mutant reveal cell-extrinsic determinants of neuronal morphology. *J. Neurosci.* **13**, 4254–4271 (1993).
33. Winkle, C. C. et al. *Trim9* deletion alters the morphogenesis of developing and adult-born hippocampal neurons and impairs spatial learning and memory. *J. Neurosci.* **36**, 4940–4958 (2016).
34. Abitua, P. B. et al. The pre-vertebrate origins of neurogenic placodes. *Nature* **524**, 462–465 (2015).
35. Abitua, P. B., Wagner, E., Navarrete, I. A. & Levine, M. Identification of a rudimentary neural crest in a non-vertebrate chordate. *Nature* **492**, 104–107 (2012).
36. Stolfi, A. et al. Early chordate origins of the vertebrate second heart field. *Science* **329**, 565–568 (2010).
37. Horie, R. et al. Shared evolutionary origin of vertebrate neural crest and cranial placodes. *Nature* **560**, 228–232 (2018).
38. Zeng, F. et al. Papillae revisited and the nature of the adhesive secreting collocytes. *Dev. Biol.* **448**, 183–198 (2019).
39. Hébert, J. M. & Fishell, G. The genetics of early telencephalon patterning: some assembly required. *Nat. Rev. Neurosci.* **9**, 678–685 (2008).
40. Zembrzycki, A., Griesel, G., Stoykova, A. & Mansouri, A. Genetic interplay between the transcription factors Sp8 and Emx2 in the patterning of the forebrain. *Neural Dev.* **2**, 8 (2007).
41. Jacquet, B. V. et al. Specification of a Foxj1-dependent lineage in the forebrain is required for embryonic-to-postnatal transition of neurogenesis in the olfactory bulb. *J. Neurosci.* **31**, 9368–9382 (2011).
42. Carlin, D. et al. Six3 cooperates with Hedgehog signaling to specify ventral telencephalon by promoting early expression of Foxg1a and repressing Wnt signaling. *Development* **139**, 2614–2624 (2012).
43. Zhong, S. et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature* **555**, 524–528 (2018).

METHODS

***Ciona* handling, collection and dissociation of embryos.** The adults of *C. intestinalis* were purchased from M-Rep. The eggs and sperm were obtained as previously described[44]. Sperm was added to the eggs for 10 min. Then, the fertilized eggs were washed twice with filtered sea water. Except in the case of the larval stage, the same animal provided both the eggs and the sperm to lower the polymorphism rate for downstream analysis. Embryos were raised to different stages at 18 °C according to a previously described method[45]. At least two biological replicates from each developmental stage were collected (Supplementary Table 1). For each sample, 100 to 500 morphologically normal embryos were randomly picked and transferred into tubes pre-coated with 5% BSA in Ca^{2+}-free artificial sea water (Ca^{2+}-free ASW, 10 mM KCl, 40 mM $MgCl_2$, 15 mM $MgSO_4$, 435 mM NaCl, 2.5 mM $NaHCO_3$, 7 mM Tris base, 13 mM Tris-HCl). Embryos were immediately dissociated with 0.5 to 1% trypsin in Ca^{2+}-free ASW with 5 mM EGTA (ASW EGTA) for 2 min (gastrula and neurula stages) to 10 min (tailbud stages). Embryos were pipetted for 5 min on ice to complete the dissociation of individual cells. Then, the digestion was inhibited with 0.2% BSA in Ca^{2+}-free ASW or quenched by 20% FBS. Cells were collected by centrifugation at 4 °C at 500*g* for 2 to 5 min and then resuspended in ice-cold Ca^{2+}-free ASW containing 0.5% BSA. For the swimming larval stage, the embryos were either homogenized (H100, Waverly) and dissociated using 1% trypsin or dissociated with 1% trypsin, 1 mg/ml collagenase, 0.5% pronase and 0.5 mg/ml cellulase in ASW EGTA. Once dissociated, the enzymes were inhibited by 20% FBS and 2 mg/ml glycine. The cells were washed and resuspended as described above.

Single-cell barcoding, on-chip and off-chip technical replicates, library preparation and sequencing. The cell concentration of each sample was checked by TC20 Automated Cell Counter to ensure it was within 1,000–2,000 cells per microlitre. Single-cell suspensions were loaded onto The Chromium Controller (10x Genomics). To assess technical variations between replicates, on-chip and off-chip experiments were performed. The on-chip experiment consisted of loading two lanes of cells from latTII embryos on the same chip, obtained by fertilizing eggs with sperm from the same animal. In the off-chip experiment, dissociated cells from latTI embryos obtained with the same fertilization strategy were loaded on the same lane on two different chips and processed separately. For all of the samples, cells were lysed, cDNAs were barcoded and amplified with Chromium Single Cell 3′ Library and Gel Bead Kit v2 (10x Genomics) following the instructions of the manufacturer. Illumina sequencing libraries were prepared from the cDNA samples using the Nextera DNA library prep kit (Illumina). All of the libraries were sequenced on Illumina HiSeq 2500 Rapid flowcells (Illumina) with paired-end 26 nucleotides (nt) + 125 nt reads following standard Illumina protocols.

Raw sequencing reads were filtered by Illumina HiSeq Control Software and only pass-filtered reads were used for further analysis. Samples were run on both lanes of a HiSeq 2500 Rapid Run mode flow cell instrument. Base calling was performed by Illumina RTA version 1.18.64.0. BCL files were then converted to FASTQ format using bcl2fastq version 1.8.4 (Illumina). Reads that aligned to phix (using Bowtie version 1.1.1) were removed, as were reads that failed Illumina's default chastity filter. We then combined the FASTQ files from each lane and separated the samples using the barcode sequences allowing one mismatch (using barcode_splitter version 0.18.2). Using 10x CellRanger version 2.0.1, the count pipeline was run with default settings on the FASTQ files to generate gene–barcode matrices for each sample. The reference sequence was obtained from the Ghost database[46].

Data quality control and visualization. To remove signals from putative empty droplet or degraded RNA, low-quality transcriptomes were filtered for each time course sample as follows: (1) we discarded cells with less than 1,000 expressed genes; (2) cells with unique molecular identifiers (UMIs) of five s.d. above the mean were not included in our analyses (Supplementary Table 1); (3) we considered only genes that were expressed in at least three cells in each dataset. In total, 90,579 cells were kept for subsequent analysis. We further normalized the read counts of each cell by Seurat methods[47], and the normalized read counts were log-transformed for downstream analyses and visualizations. For dimensional reduction, the relative expression measurement of each gene was used to remove unwanted variation. Genes with the top 1,000 highest standard deviations were obtained as highly variable genes. After significant (jackstraw procedure) principal components were identified, a graph-based clustering approach was used for partitioning the cellular distance matrix into clusters. Cell distance was visualized by *t*-SNE method in reduced two-dimensional space.

For *t*-SNE visualization of the whole dataset (as shown in Fig. 1b, Extended Data Fig. 2a, b), the shared highly variable genes (500 genes) among all of the samples were kept for canonical correlation analysis. The top 50 canonical correlation vectors were calculated, and each dataset was projected into the maximally correlated subspaces. According to the relationship between the number of canonical correlation vectors and the percentage of variance explained, 1–18 canonical correlation vectors were used for subspace alignment[47] and dimensional reduction. Graph-based clustering was performed in the lower-dimensional space, and an approximate nearest neighbour search was performed. We used ten random starts for clustering and selected the result with highest modularity. A modified fast Fourier transform-accelerated interpolation-based *t*-SNE[48] was used in the visualization of our dataset. Maximum iteration times were set to 2,000 and the perplexity parameter was set to 30.

For *t*-SNE visualization of cells in each tissue type from different developmental stages, we regressed out the effects produced by the UMI counts, experiment batch and sample identities with negative binomial regression modelling before dimensional reduction and clustering. Similar to approaches mentioned above, after significant (jackstraw procedure) principal components were identified (1–20 principal components were used for the subclustering of cells in each tissue type), graph-based clustering was performed and a modified fast Fourier transform-accelerated interpolation-based *t*-SNE was used in the visualization.

Annotation of cell clusters. We annotated the clusters for assigning clustering results to tissue types. For the clustering applied to *t*-SNE coordinates of the whole dataset, three steps were followed to refine the annotation results. (1) The expression pattern of top marker genes and regulatory genes were compared between clusters. Clusters with similar expression pattern of key regulatory genes and known markers were considered to be the same tissue type. (2) We carefully compared the annotation results with the in situ images recorded in the Ghost (http://ghost.zool.kyoto-u.ac.jp) and Aniseed (https://www.aniseed.cnrs.fr/aniseed/experiment/find_insitu) databases, or published papers, to validate the annotation results. (3) For putative newly discovered cell types in clusters with poorly annotated marker genes, we carefully checked the gene-expression pattern to make sure there was no ambiguous expression of known markers, which might indicate cell doublets. We also consulted with experts in ascidian research to validate our findings. The newly identified cell types were named according to their specifically expressed genes—for example, $Tll1^+$ cells in the mesenchyme. If the position and morphology of the cells were verified by our reporter assays, we compared them to published papers with morphological information (for example, the previously published synaptome[8]) and identified the cell types (for example, Eminens neurons).

Cell-state mapping across time points. Most of the major tissues of the ascidian tadpole are first specified before gastrulation, which is our starting time point for sampling. To capture the developmental transitions that stem from different blastomeres at the 110-cell stage, we performed 'ancestor voting' between clusters across time, as previously described[1]. In brief, between every two adjacent time points, all of the cells were embedded into the principal component analysis space of the later time point (1–50 principal components were used). For each cluster identified in the later time point, each cell in the cluster and its nearest five neighbouring cells in the previous time point were calculated based on the similarity between transcriptomes. The voting results for all of the cells in each cluster of the later time point were aggregated, and the percentage of 'ancestor cells' in each cluster of the earlier time point was calculated. In cases in which a cluster had more than one ancestor cluster, the cluster in the earlier time point was considered to be the winning ancestor if it had a percentage number that was at least two times that of the other clusters.

To better capture all of the subpopulations for highly differentiated tissue types, we subclustered the cells from each of the stages and performed the ancestor voting process between subclusters across time points. For mesenchyme cells, most of the winning ancestors were unambiguous assigned, with a >90% winning share on average. For cells from the epidermis and nervous system systems (as some sensory neurons differentiate from the epidermis), we subclustered the epidermis and nervous system cells together. We deleted the subclusters if they were assigned to multiple ancestor clusters and there was no winning ancestor (no winning cluster met our criteria (that is, a percentage number that was at least twice of the other clusters)) to make sure the intermediate state or immature neuron types did not affect the cell-state mapping results.

Single-cell trajectory construction. For notochord and Eminens neurons, we used monocle 2[49] to construct the single-cell trajectory for each cleanly defined lineage with known markers. Highly variable genes with $q < 0.01$ were selected across time points, a discriminative dimensionality reduction (DDRTree) was performed with regression on the UMI counts to eliminate unwanted variation introduced by sequencing depth between samples, and cells were ordered along the trajectory according to their pseudotime value. A subset of significant genes with $q < 1 \times 10^{-100}$ and $q < 1 \times 10^{-20}$ are shown in the pseudotemporal expression pattern of primary and secondary notochord, respectively, in Extended Data Fig. 5.

For tissues that contained more complexity during development, such as the mesenchyme and nervous system (which had 9 and 41 identified subclusters at the larval stage, respectively), we used a simulated diffusion-based computational reconstruction method (URD[2]) for acquiring the transcriptional trajectories during embryogenesis. In brief, after differentially expressed genes were picked and dimensional reduction was performed for cells in specific tissues from ten developmental stages as described in 'Data quality control and visualization', we calculated

transition probabilities between cells with the destiny package[50]. We next assigned to the cells a pseudotime value with a probabilistic breadth-first graph search using the transition probabilities. To find the developmental trajectories, we performed biased random walks that started from a random cell in each refined cluster of the final stage (that is, the larval stage) that we covered. The walk was simulated through cells on the basis of the transition probabilities, and the transitions were only allowed for cells with younger or similar pseudotimes to make sure the trajectory between the root (cells from the earliest time point) and the tip (cells from the last time point) was found. Then, the biased random walk was processed into visitation frequencies. The URD tree structure was built by aggregating trajectories when the same cells were visited from each tip.

For cells in the mesenchyme, we optimized the number of nearest neighbours (*k*-nearest neighbour) and set it to 250, and the width of the Gaussian used to transform cell–cell distances into transition probabilities (sigma) was set to six. We also modified parameters for constructing the URD tree as follows: divergence.method = "preference", cells.per.pseudotime.bin = 75, bins.per.pseudotime.window = 10, p.thresh = 0.01. To avoid ambiguities in reconstructing gene-expression lineages for cells in the nervous system, we excluded or combined those cell clusters that (1) were not well-defined or determined during neurogenesis on the basis of prior knowledge; (2) could not be resolved by diffusion components (such as very small population of cells; for example, decussating neurons); and (3) exhibited intermixing in the diffusion maps. The parameters were set as follows: divergence.method = "preference", cells.per.pseudotime.bin = 28, bins.per.pseudotime.window = 4, p.thresh = 0.025, minimum.visits = 40. For endodermal cells, the parameters were set as follows: divergence.method = "preference", cells.per.pseudotime.bin = 65, bins.per.pseudotime.window = 10, p.thresh = 0.01, minimum.visits = 20. For muscle cells, the parameters were set as follows: divergence.method = "preference", cells.per.pseudotime.bin = 20, bins.per.pseudotime.window = 10, p.thresh = 0.01, minimum.visits = 20.

Gene-expression cascade. The genes included in the cascade of each trajectory were recovered following the criteria set in the URD package: cells in the segment were compared in a pairwise manner with cells from each of that segment's siblings and children, and differentially expressed genes were kept if they were expressed in more than 10% of the population, their mean expression level was 1.5× higher than in the sibling branch, and the genes were 1.25× better classifiers than a random classifier for the population. Then, an impulse model was fitted to the expression of each gene recovered in the cascade for determining the 'on and off' timing of expression, and the genes were ordered by the 'on-time' in the cascade. Genes with an expression pattern that was not fitted with the impulse model were arranged at the bottom of the cascade. In the heat map, cells were ordered with the progression of pseudotime using a moving window, and the scaled mean expression within each pseudotime moving-window was plotted.

Regulatory network. Regulatory genes, signalling pathway genes recovered in each developmental trajectory cascade and selected highly expressed genes for specific cell types at the final time point that had a fold change (expressed in $\log_2$) above one between groups were all used in investigating the putative direct interaction. We used cluster-buster[51] to find clusters of pre-specified motifs 2 kb upstream of the transcription start site of each gene. The parameters were set as follows: $g = 1$, $m = 0$, $c = 0$, score ≥ 6. The position frequency matrix was downloaded from the JASPAR 2018 database[52]. Genes with no position frequency matrix recorded in JASPAR was not considered in constructing the regulatory network. The regulatory network was plotted with Biotapestry. Each line between every two genes represents a putative direct interaction, as the binding motif of the regulatory gene was identified in the motif-cluster region of the target gene.

Heat maps. Heat maps in Extended Data Fig. 3 were plotted with the DoHeatmap function of Seurat v.2.3.2. Only genes with an average fold change (expressed logarithmically) > 0.3 are shown. For Extended Data Fig. 5d, differentially expressed genes between primary notochord and secondary notochord were identified by the following criteria using DESeq2[53]: (1) FDR (false discovery rate) adjusted *P* value below 0.05; and (2) absolute fold change (expressed in $\log_2$) between groups was larger than 1.5. The mean expression level of each gene within one developmental stage was calculated, and the scaled expression of the genes was on the basis of the Euclidean distance using pheatmap 1.0.10. For Fig. 3e, genes with an average fold change (expressed logarithmically) > 1.5 are shown. Both Fig. 3e and Extended Data Fig. 5d were plotted with pheatmap. The pseudotemporal expression heat maps in Extended Data Fig. 5b, c and Extended Data Fig. 9a, and the expression dynamics in Fig. 3b, were plotted using monocle 2.

Molecular cloning. The KH number of all of the genes mentioned in the manuscript as well as other names that are commonly used in the *Ciona* field can be found in the Supplementary Table 6.

Dmbx, *Dmrt1*, *Gad*, *Prop*, *Twist* and *vGat* regulatory sequences have previously been described[27,31,34,54,55]. They were cloned in *pCESA* expression vector upstream of the reporter genes *GFPCAAX* (CAAX is the palmitoylation motif to target a protein to the membrane), *H2B-mApple*, *H2B-YFP*, *mNeonGreen-PH* (*nG-PH*), *mCherryCAAX* and *H2B-mCherry* using NotI and AscI restriction enzymes (NEB). The expression vector with *H2B-mApple* reporter construct was obtained by inserting *mApple*[56] (primers in Supplementary Table 7) into the *pCESA* expression vector that contains *H2B*, using NEBuilder (NEB). The expression vector that contains the *nG-PH* reporter gene was obtained by first inserting *GFP-PH* (courtesy of T. Meyer)[57] using NotI and FseI (NEB) into a *pCESA* expression vector and then replacing the *GFP* coding sequence with *mNeonGreen*[58] by recombination using NEBuilder (primers in Supplementary Table 7).

Asic1b, *Calm*, *Fgf13*, *Galr2*, *S39aa*, *S39aa* 2.2 kb and *Znt3* regulatory sequences were PCR-amplified (primers in Supplementary Table 7) from genomic DNA and cloned into *pCESA-H2B:mCherry* using AscI and NotI restriction enzymes.

After PCR amplification (primers in Supplementary Table 7) *Casq1/2* regulatory sequences were cloned into an expression vector that contains *GFP* downstream of the minimal promoter of *fog* (*pCESA-fog>GFP*) using AscI and XbaI restriction enzymes (NEB). The regulatory sequences of *NP* (*KH.C11.631*) were PCR-amplified and cloned into *pCESA-fog>GFPCAAX*.

After PCR amplification from the *Prop>GFPCAAX* (primers in Supplementary Table 7), *Prop* 900 bp, *Prop* 700 bp and *Prop* 300 bp were cloned into *pCESA-GFPCAAX* vector using AscI and NotI.

For live imaging, *Prop* 700 bp was also cloned upstream of *PH-nG*. The reporter gene was obtained by NEBuilder assembly. First the *PH* domain, *GFP* and the degradation signal of *Hes-b* (*PH-GFP* primers), which was obtained from *Ciona* cDNA, were assembled into a *pCESA* expression vector using NEB builder. Then, *GFP* and the degradation signal coding sequences were replaced by *mNeonGreen* and a shorter degradation sequence using NEBuilder assembly (*PH-nG* primers). Finally a second degradation signal was inserted before the stop codon using NEBuilder assembly (deg primers, primers in Supplementary Table 7).

Prop 260 bp was cloned into *pCESA-fog GFPCAAX* vector using AscI and XbaI. Point mutations in FoxH-a binding site of the *Prop* 260 bp regulatory sequences were obtained by plasmid PCR of *Prop* 260 bp *fog>GFPCAAX* (primers in Supplementary Table 7).

Galr2 regulatory sequences specifically active in bipolar tail neurons were amplified from *Galr2>H2B-mCherry* (primers in Supplementary Table 7) and cloned into *pCESA-fog>mCherryCAAX* using AscI and XbaI restriction sites.

Tll1, *Hlx* and *FoxG* regulatory sequences were PCR-amplified (primers in Supplementary Table 7) and then assembled into *pSP-Kaede* expression vector using NEBuilder. *Ptf1a* was obtained by PCR-amplifying an expression vector that contains the full-length *Ptf1a* regulatory sequences[21] (primers in Supplementary Table 7). The PCR product was self-recombined using NEBuilder. *Ptf1a* was then subcloned upstream of *mCherryCAAX* in the *pCESA* expression vector.

A *LacZ* expression vector under the control of *Dmrt1* (*Dmrt1>LacZ*) has previously been described[54]. The *Prop* coding sequence was amplified from mid-tailbud embryo cDNA and cloned downstream of *Dmrt1* regulatory sequences using NotI and FseI restriction enzymes (NEB).

***Ciona* electroporation and imaging.** After fertilization, one-cell-stage embryos were electroporated using 20 to 100 μg of each expression construct as previously described[10].

The embryos were raised at 16 °C, 18 °C or 21 °C in ASW and fixed at the desired stage following a previously described protocol[54]. The embryos were washed several times with 0.05% BSA in PBS before being mounted using FluorSave Reagent (Millipore). Images were acquired with a Zeiss 880 confocal microscope with or without the Airyscan module, and a wide-field Zeiss Axio Observer Z1/7 combined to the Apotome 2.0 module.

All electroporation was performed in duplicate or triplicate. Between 18 and 610 embryos were recovered per condition. No specific randomization strategy was performed, except for the assignment of the fertilized eggs to the different conditions.

Live imaging was performed using a two-photon microscope system built in-house. Embryos were anaesthetized with 16 mg/ml MS-222 in ASW (Sigma-Aldrich). They were placed in microwells cast in 1% agarose in ASW[59], and the imaging was performed at 18 °C from the latTI to latTIII stage. The images were assembled using Fiji[60] and the final rendering obtained with Imaris (Bitplane).

Statistical analysis of the functional assays. For the statistical tests, the embryos with the same electroporated plasmids were pooled over the different experiments. Mann–Whitney *U*-test was performed with the package Tidyverse of R software, the χ^2 test followed by the post hoc test for pairwise comparison, Fisher's test with Bonferroni adjustment was also performed with Tidyverse[61].

Fish husbandry, generation of transgenic fish and imaging. All experiments with the African killifish *N. furzeri* were performed using the GRZ strain. All of the fish were housed at 27 °C in a facility overseen by the Stowers Institute for Medical Research (SIMR) Institutional Animal Care and Use Committee. Work with fish was performed according to the guidelines of the Stowers Institute for Medical Research.

A 4-kb *Ciona FoxG* regulatory sequence was cloned into *pDest-Tol2-miniP-GFP-Cryaa-Venus* transgenic vector through Gibson assembly. To generate the

transgenic killifish, 15–20 pg DNA was co-injected with 30 pg transposase mRNA into one-cell-stage *N. furzeri* embryos and the injected embryos were maintained in Yamamoto embryo solution (17 mM NaCl, 2.7 mM KCl, 2.5 mM $CaCl_2$, 0.02 mM $NaHCO_3$, pH 7.3) at 28 °C for 2 weeks before hatching. F_0 founders were crossed with wild-type GRZ fish and three independent lines were established for gene expression studies. No genotyping was performed to detect the transgene. However, for the first transgenic line, 15 out of 46 F_1 embryos showed GFP expression in the forebrain. For the second transgenic line, 8 out of 25 F_1 embryos showed GFP expression in the forebrain. Finally, 10 out 37 F_1 embryos of the third transgenic line had GFP expression in the forebrain. No particular randomization strategy was implemented.

Killifish embryos were removed manually from the chorion before imaging. The juvenile fish were anaesthetized in 150 mg/l MS-222 for 5 min at room temperature. Images were taken with Ultraview R2 spinning disk confocal microscope.

Estimation of sample size, blinding and randomization. No statistical methods were used to predetermine sample size. For the single-cell experiments, because the embryo collection and the subsequent data analysis was performed by different researchers, the investigators were blinded to group allocation. For the functional assays, no particular blinding strategy was adopted. As stated in the specific sections above, the assignment of the *Ciona* embryos to the different conditions were randomized. Otherwise, no particular randomization strategy was used.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw sequencing data and the gene-expression matrix are available in the Gene Expression Omnibus (GEO) under accession number GSE131155. Our data can be explored at https://portals.broadinstitute.org/single_cell/study/SCP454/comprehensive-single-cell-transcriptome-lineages-of-a-proto-vertebrate. All other data are available from the corresponding authors on reasonable request.

44. Christiaen, L., Wagner, E., Shi, W. & Levine, M. Isolation of sea squirt (*Ciona*) gametes, fertilization, dechorionation, and development. *Cold Spring Harb. Protoc.* **2009**, pdb.prot5344, (2009).
45. Hotta, K. et al. A web-based interactive developmental table for the ascidian *Ciona intestinalis*, including 3D real-image embryo reconstructions: I. From fertilized egg to hatching larva. *Dev. Dyn.* **236**, 1790–1805 (2007).
46. Satou, Y., Kawashima, T., Shoguchi, E., Nakayama, A. & Satoh, N. An integrated database of the ascidian, *Ciona intestinalis*: towards functional genomics. *Zool. Sci.* **22**, 837–843 (2005).
47. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
48. Linderman, G. C., Rachh, M., Hoskins, J. G., Steinerberger, S. & Kluger, Y. Fast interpolation-based *t*-SNE for improved visualization of single-cell RNA-seq data. *Nat. Methods* **16**, 243–245 (2019).
49. Qiu, X. et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017).
50. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).
51. Frith, M. C., Li, M. C. & Weng, Z. Cluster-Buster: finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res.* **31**, 3666–3668 (2003).
52. Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* **46**, D1284 (2018).
53. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
54. Wagner, E. & Levine, M. FGF signaling establishes the anterior border of the *Ciona* neural tube. *Development* **139**, 2351–2359 (2012).
55. Yoshida, R. et al. Identification of neuron-specific promoters in *Ciona intestinalis*. *Genesis* **39**, 130–140 (2004).
56. Shaner, N. C. et al. Improving the photostability of bright monomeric orange and red fluorescent proteins. *Nat. Methods* **5**, 545–551 (2008).
57. Stauffer, T. P., Ahn, S. & Meyer, T. Receptor-induced transient reduction in plasma membrane PtdIns(4,5)P2 concentration monitored in living cells. *Curr. Biol.* **8**, 343–346 (1998).
58. Shaner, N. C. et al. A bright monomeric green fluorescent protein derived from *Branchiostoma lanceolatum*. *Nat. Methods* **10**, 407–409 (2013).
59. Gregory, C. & Veeman, M. 3D-printed microwell arrays for *Ciona* microinjection and timelapse imaging. *PLoS ONE* **8**, e82307 (2013).
60. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
61. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2013).

Acknowledgements We thank J. B. Wiggins, J. M. Miller and the Genomics Core Facility for technical support of 10x Chromium platform; IT and Bioinformatics staff at the Lewis-Sigler Institute for Integrative Genomics (LSI) for development of the sequence alignment pipeline; E. G. Gatzogiannis (director of the LSI Imaging Core Facility) for building the two-photon microscope and help with imaging; the Shvartsman laboratory in LSI for Imaris access; A. Sánchez Alvarado at SIMR for providing laboratory support and resources; and members of the Levine Laboratory for helpful discussions, and especially N. Treen for suggesting mNeonGreen as fluorescent reporter. This study was supported by a grant from the NIH (NS076542) to M.L. W.W. (SIMR) was funded by the Stowers Institute.

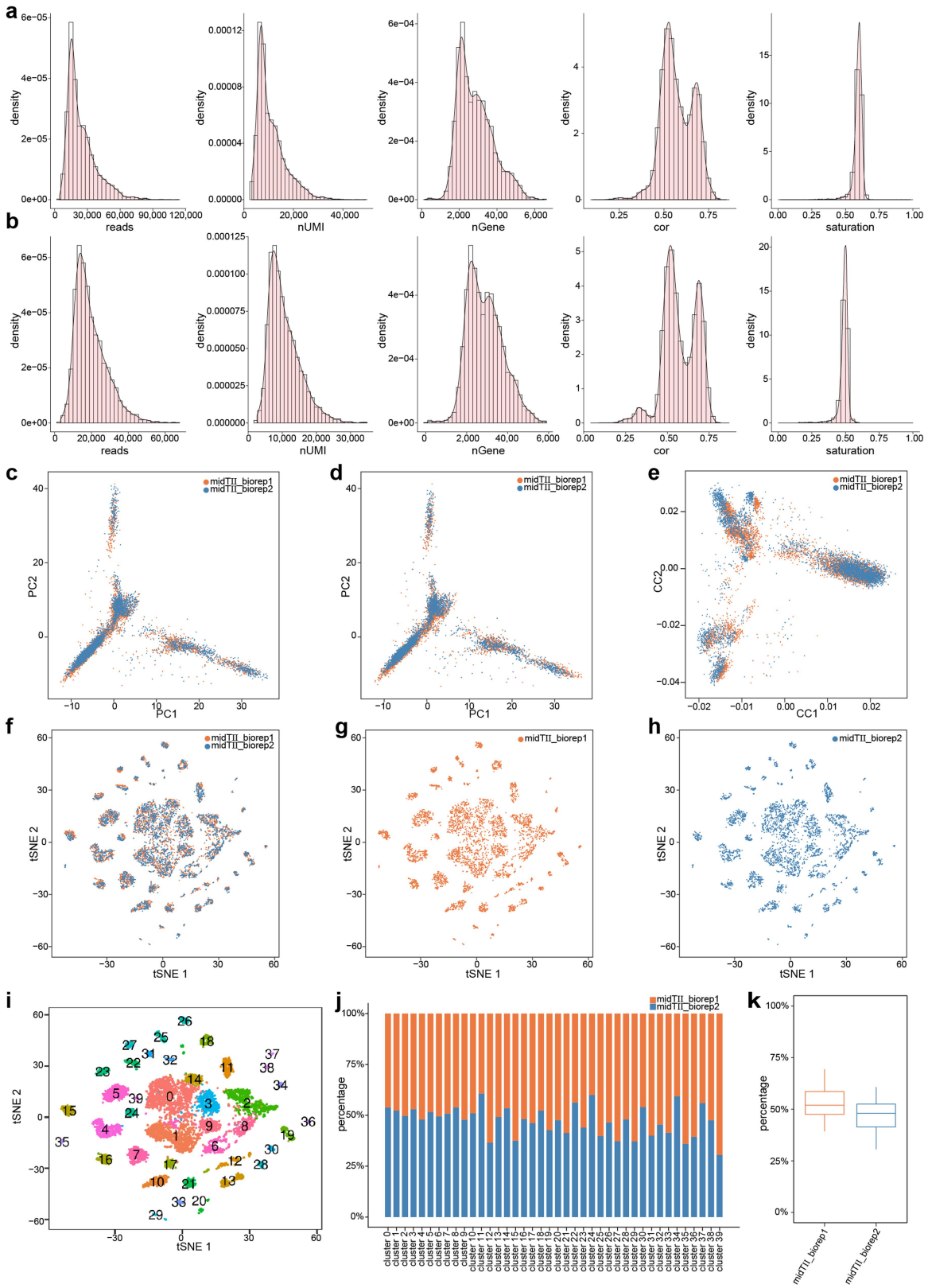
Author contributions K.C. and M.L. conceived the project. K.C., M.L., C.C., L.A.L. and W.W. (SIMR) designed the experiments. L.A.L., P.H.Y., Y.A.C. and K.C. performed *Ciona* experiments, W.W. (SIMR) performed killifish experiments, C.C. performed computational data analysis. L.R.P., J.C.M. and W.W. (LSI) set up the single-cell RNA-sequencing pipeline. M.L. supervised the project. All authors contributed to interpretation of the results, and C.C., L.A.L., K.C. and M.L. wrote the manuscript.

Competing interests The authors declare no competing interests.

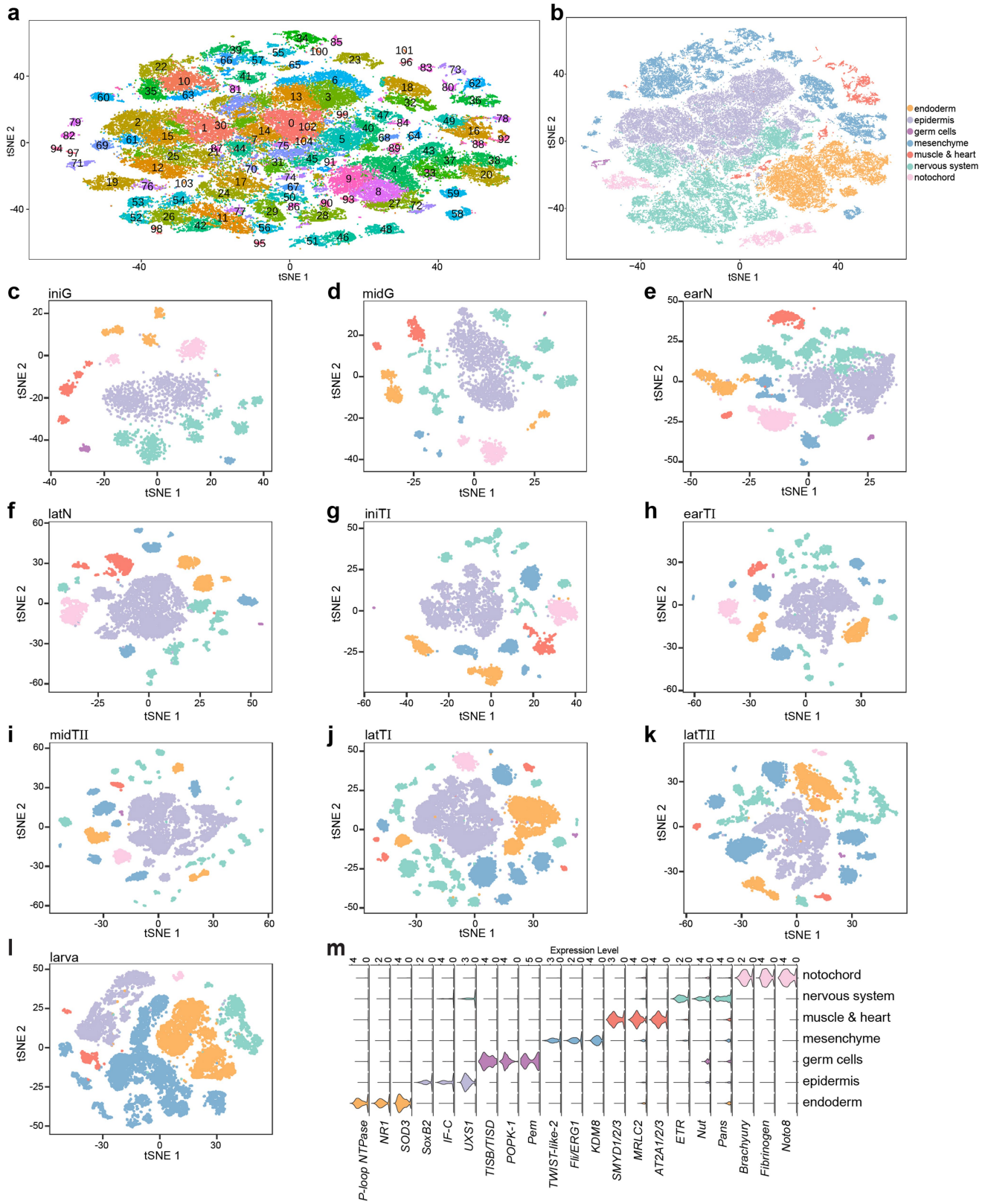
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1385-y.

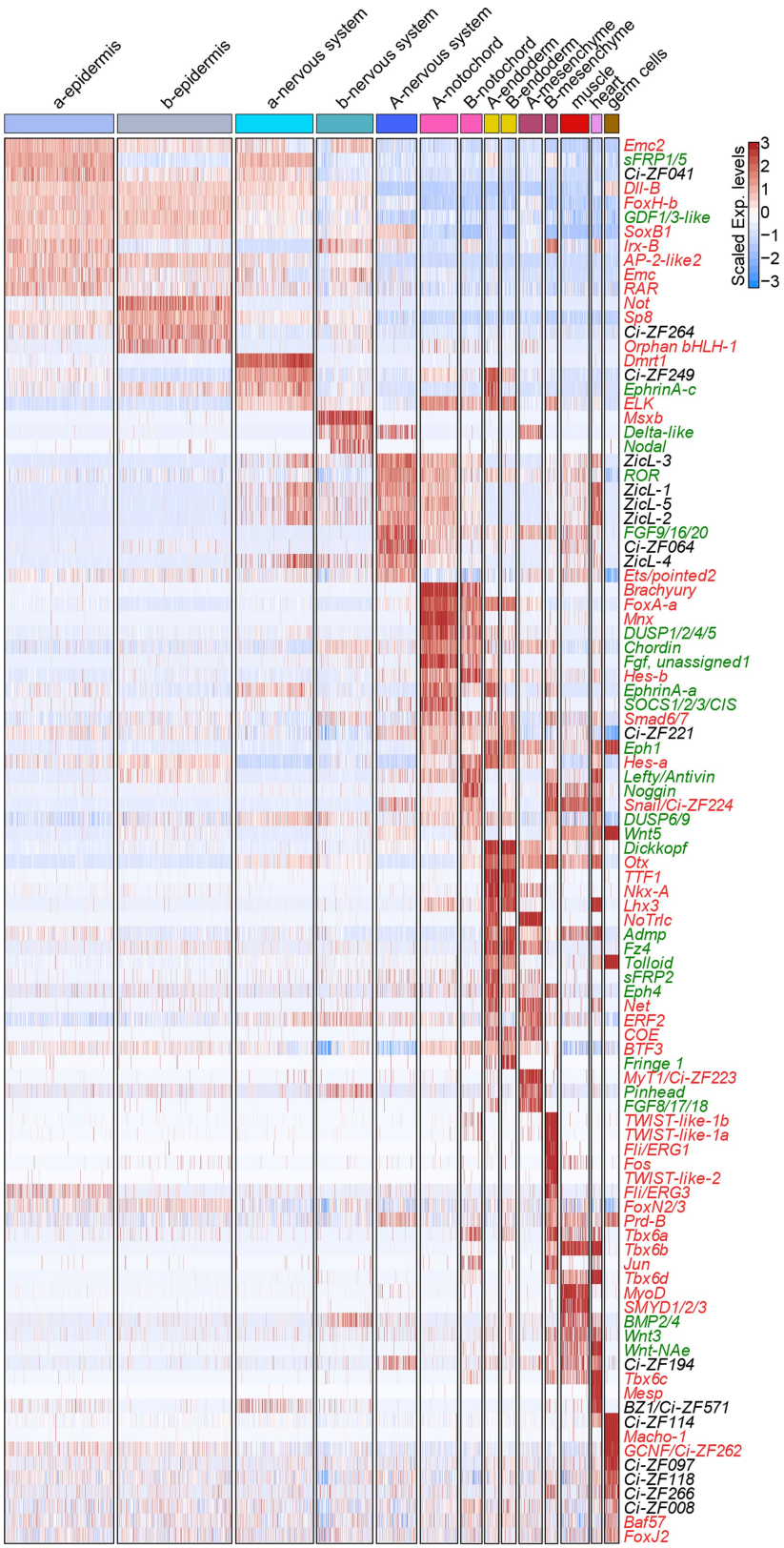
Correspondence and requests for materials should be addressed to M.L. or K.C.



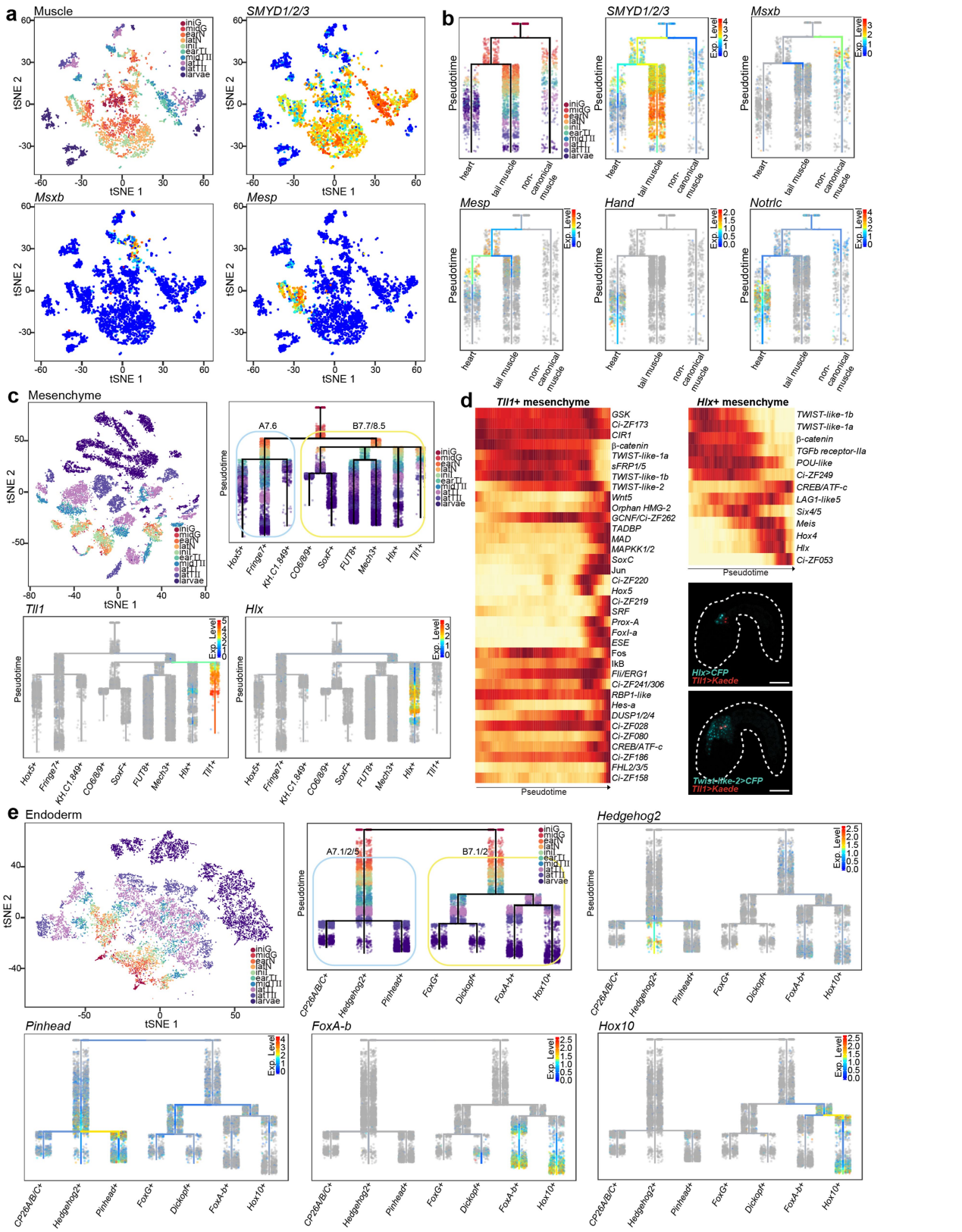
Extended Data Fig. 1 | Data quality and biological replicates from mid-tailbud stage. a, **b**, Distribution plot of reads numbers, UMIs, gene numbers, correlation coefficient (Spearman) and saturation level per cell from mid-tailbud (**a**, midTII_biorep1; **b**, midTII_biorep2). **c**, The first two principal components were plotted for cells regressed by UMIs (midTII_biorep1, $n = 4{,}929$ cells; midTII_biorep2, $n = 4{,}062$ cells). **d**, The first two principal components were plotted for cells regressed by both UMIs and batches. **e**, The first two canonical correlation vectors were plotted after alignment by canonical correlation analysis. **f**–**h**, Merged (**f**) and split (**g**, **h**) *t*-SNE clustering for the biological replicates. **i**, *t*-SNE plot of canonical-correlation-analysis-aligned samples of biological replicates ($n = 8{,}991$ cells). The numbers indicate different clusters. **j**, The percentage of cells between replicates within the same cluster (clusters shown in **i**). **k**, Box plot of the percentage of cells in each cluster ($n = 40$ clusters) between replicates. The lower, middle and upper hinges correspond to the first and third quartiles (the 25th and 75th percentiles), and the middle hinge corresponds to the median.



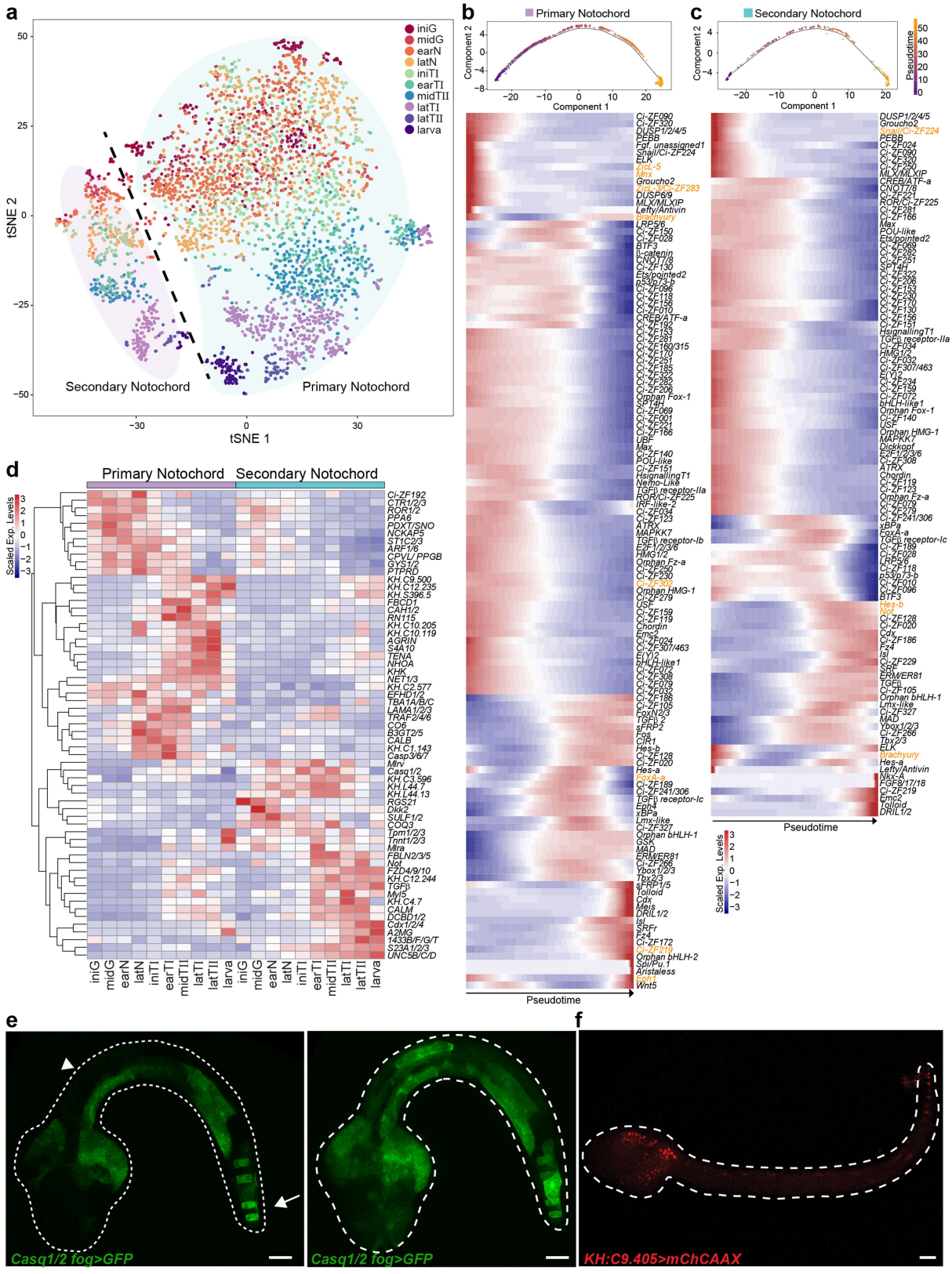
Extended Data Fig. 2 | *t*-SNE projections of ten stages from single-cell RNA-sequencing data. a, *t*-SNE plot of the entire dataset ($n = 90{,}579$ cells). Cells are coloured and labelled by clusters. Differentially expressed genes in each cluster can be found in Supplementary Table 2. **b**, *t*-SNE plot of all of the cells, coloured according to tissue type. **c**–**l**, *t*-SNE projections of cells, coloured by tissue types at different stages of development (iniG, $n = 2{,}863$ cells; midG, $n = 3{,}384$ cells; earN, $n = 7{,}154$ cells; latN, $n = 8{,}449$ cells; iniTI, $n = 5{,}668$ cells; earTI, $n = 7{,}109$ cells; midTII, $n = 8{,}991$ cells; latTI, $n = 18{,}535$ cells; latTII, $n = 12{,}635$ cells; and larva, $n = 15{,}791$ cells). The colour code is the same as in **b**. **m**, Violin plots illustrating expression levels of representative marker genes per cell per tissue type (endoderm, $n = 14{,}162$ cells; epidermis, $n = 26{,}936$ cells; germ cells, $n = 396$ cells; mesenchyme, $n = 19{,}143$ cells; muscle and heart, $n = 3{,}691$ cells; nervous system, $n = 22{,}198$ cells; and notochord, $n = 4{,}053$ cells). Colour code is the same as in **b**.



Extended Data Fig. 3 | Specification of cell types at the onset of gastrulation. The heat map shows the scaled expression of differentially expressed genes that encode transcription factors (red) and cell-signalling components (green). Many marker genes were newly identified for each tissue.

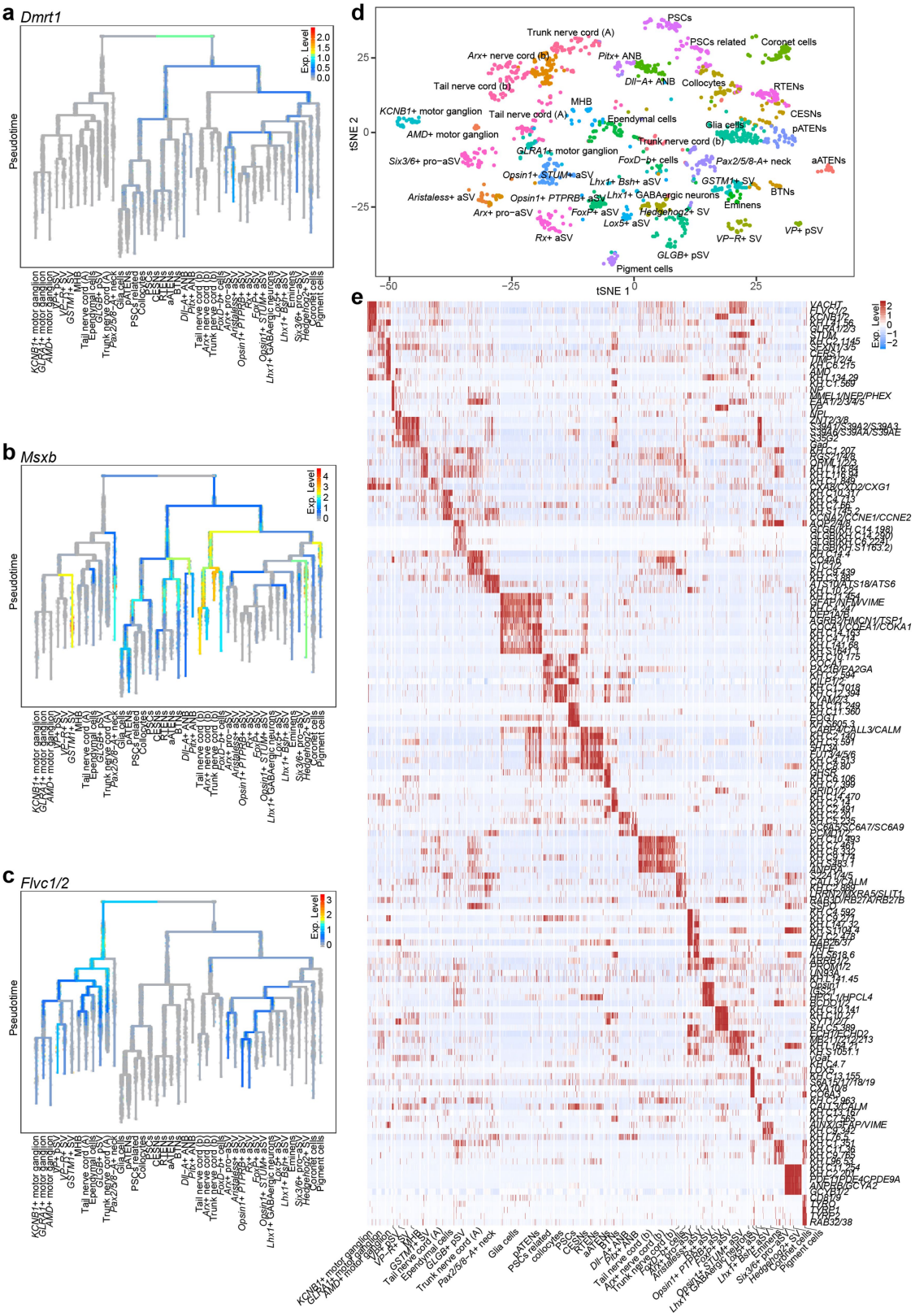


Extended Data Fig. 4 | Reconstructed transcriptional trajectories of muscle, mesenchyme and endoderm. a, *t*-SNE projection and expression patterns of representative marker genes of tail muscle, non-canonical muscle and heart ($n = 3{,}691$ cells). **b**, Reconstructed transcriptome trajectories and expression patterns of representative marker genes in muscle. **c**, *t*-SNE projection and expression patterns of representative marker genes shown on reconstructed transcriptome trajectories of mesenchyme ($n = 19{,}143$ cells). **d**, Cascade of representative transcription factors and signalling pathway genes along pseudotime in *Tll1*$^+$ and *Hlx*$^+$ mesenchyme. Mid-tailbud embryos that express *Twist-like-2* (cyan), a mesenchymal marker, and *Tll1* (red) reporter gene (top), and an *Hlx* (cyan) and *Tll1* (red) reporter gene (bottom, $n = 3$ electroporation experiments). **e**, *t*-SNE projection and expression patterns of representative marker genes shown on seven reconstructed transcriptome trajectories of endoderm ($n = 14{,}162$ cells). Scale bars, 50 μm.

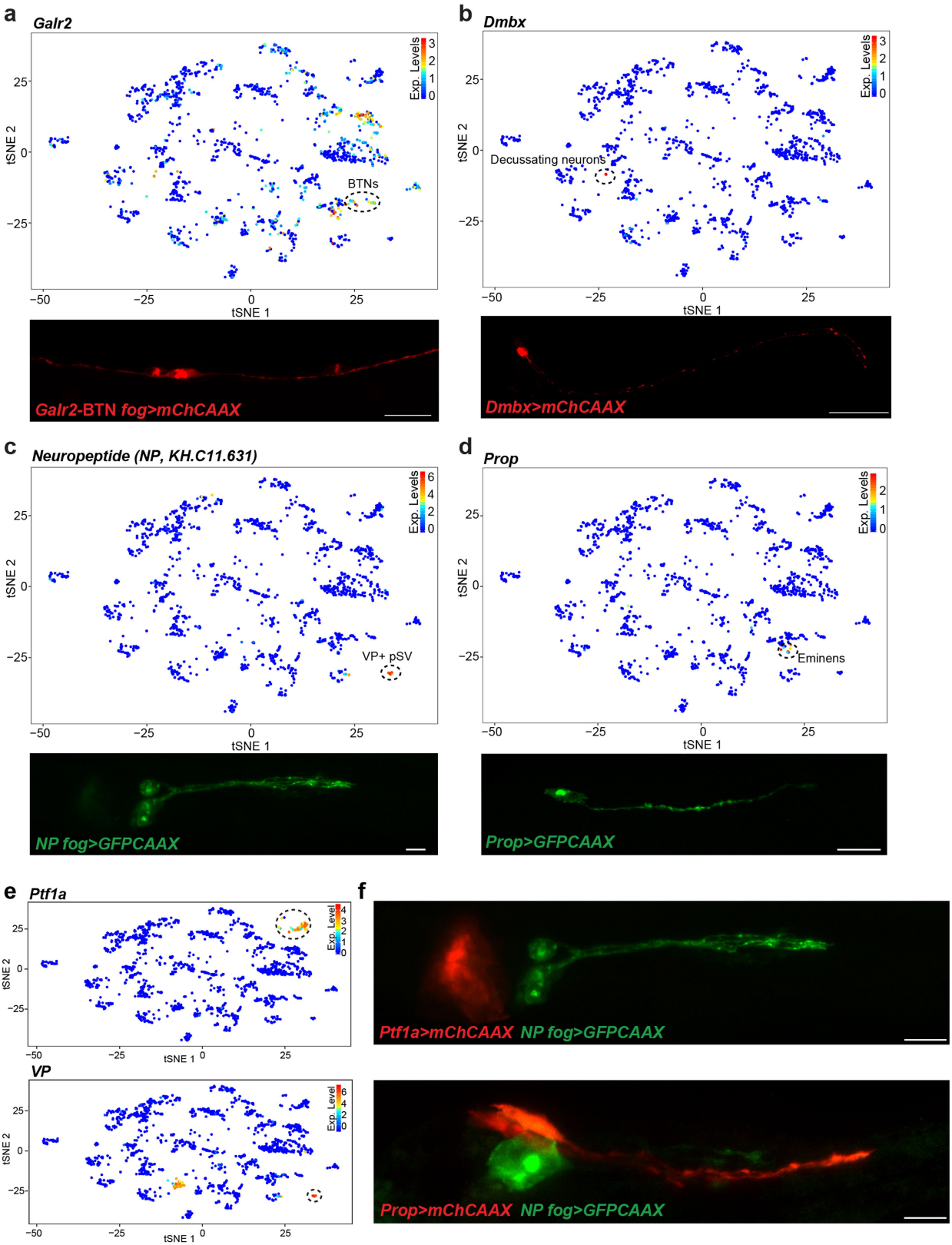


Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Transcriptome profiles of *Ciona* notochord cells during development. a, *t*-SNE plot of notochord cells. Cells are coloured by developmental stage. The dashed line shows the separation between the primary ($n = 3{,}123$ cells) and secondary lineages ($n = 627$ cells). **b**, **c**, The single-cell transcriptome trajectory (top) and pseudotemporal gene-expression profiles (bottom) of the primary notochord and the secondary notochord. Cells were ordered along the trajectory across pseudotime. Only significantly expressed genes (likelihood ratio test) with $q < 1 \times 10^{-100}$ (primary notochord) and $q < 1 \times 10^{-20}$ (secondary notochord) are shown. Selected transcription factors and signalling molecules are labelled in orange. **d**, Heat map of differentially expressed genes between the primary and secondary notochord. Genes are clustered by Euclidean distance. **e**, Expression of a *Casq1/2 fog>GFP* reporter gene in a late-tailbud-stage embryo (left, one optical plane; right, maximum intensity projection). $n = 3$ electroporation experiments. GFP (green) was present in the muscle and in the secondary notochord (arrow), but no expression was observed in the primary notochord (arrowhead). **f**, Expression of *KH.C9.405>mChCAAX* reporter gene in late tailbud II stage embryo. *mChCAAX* (red) was present in the secondary notochord but not the primary notochord. $n = 3$ electroporation experiments. Scale bars, 20 μm.

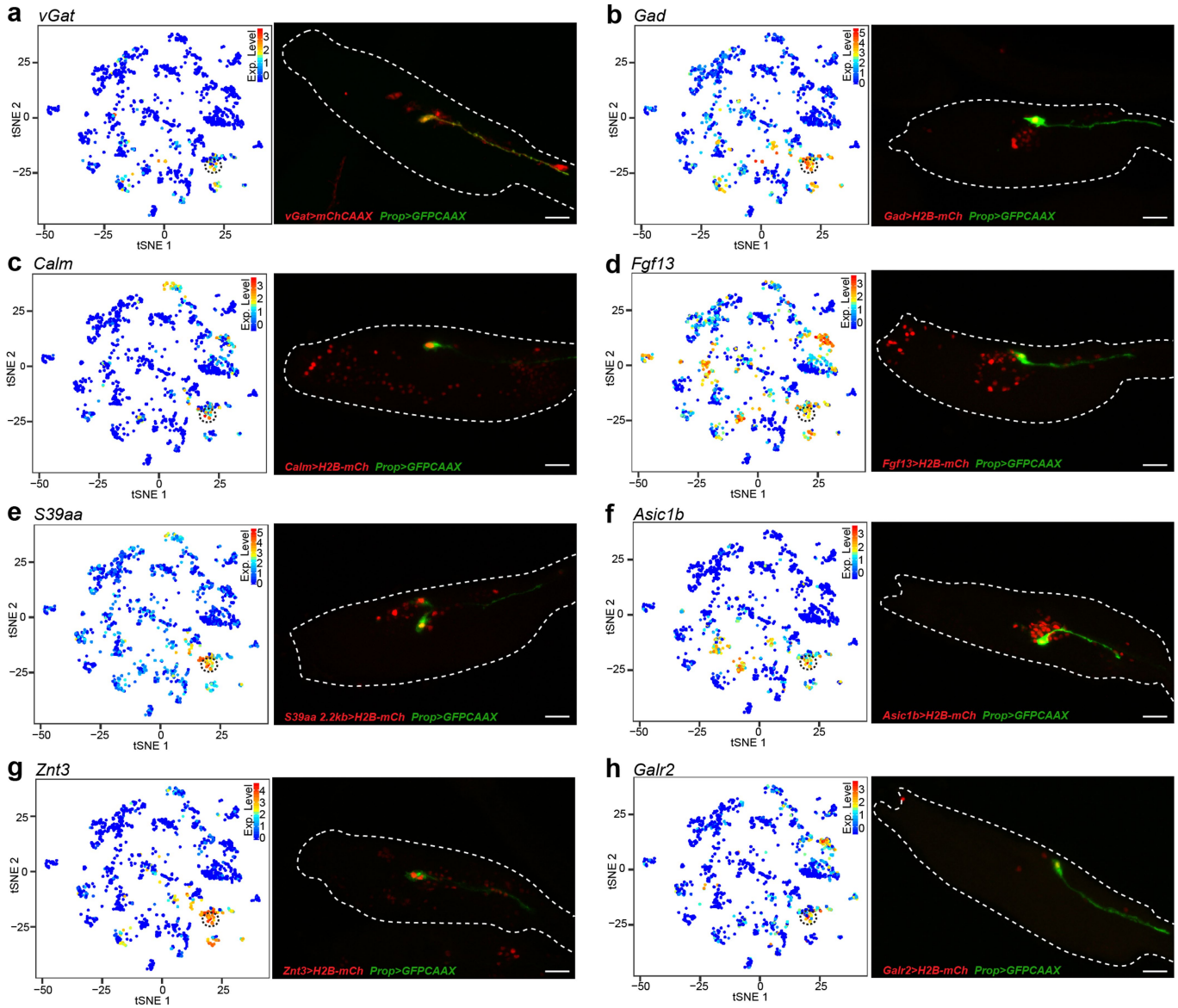


Extended Data Fig. 6 | Neural cells. **a**–**c**, Expression patterns of representative marker genes for the a- (**a**), b- (**b**) and A-lineages (**c**) are shown in reconstructed transcriptome trajectories of neural cells that span ten developmental stages. **d**, *t*-SNE plot of neural cells recovered from the larval stage ($n = 1{,}704$ cells). Identified cell types are labelled. **e**, Heat map of the top-five differentially expressed genes (not including those encoding transcription factors) for each type of neural cell in the larval stage.

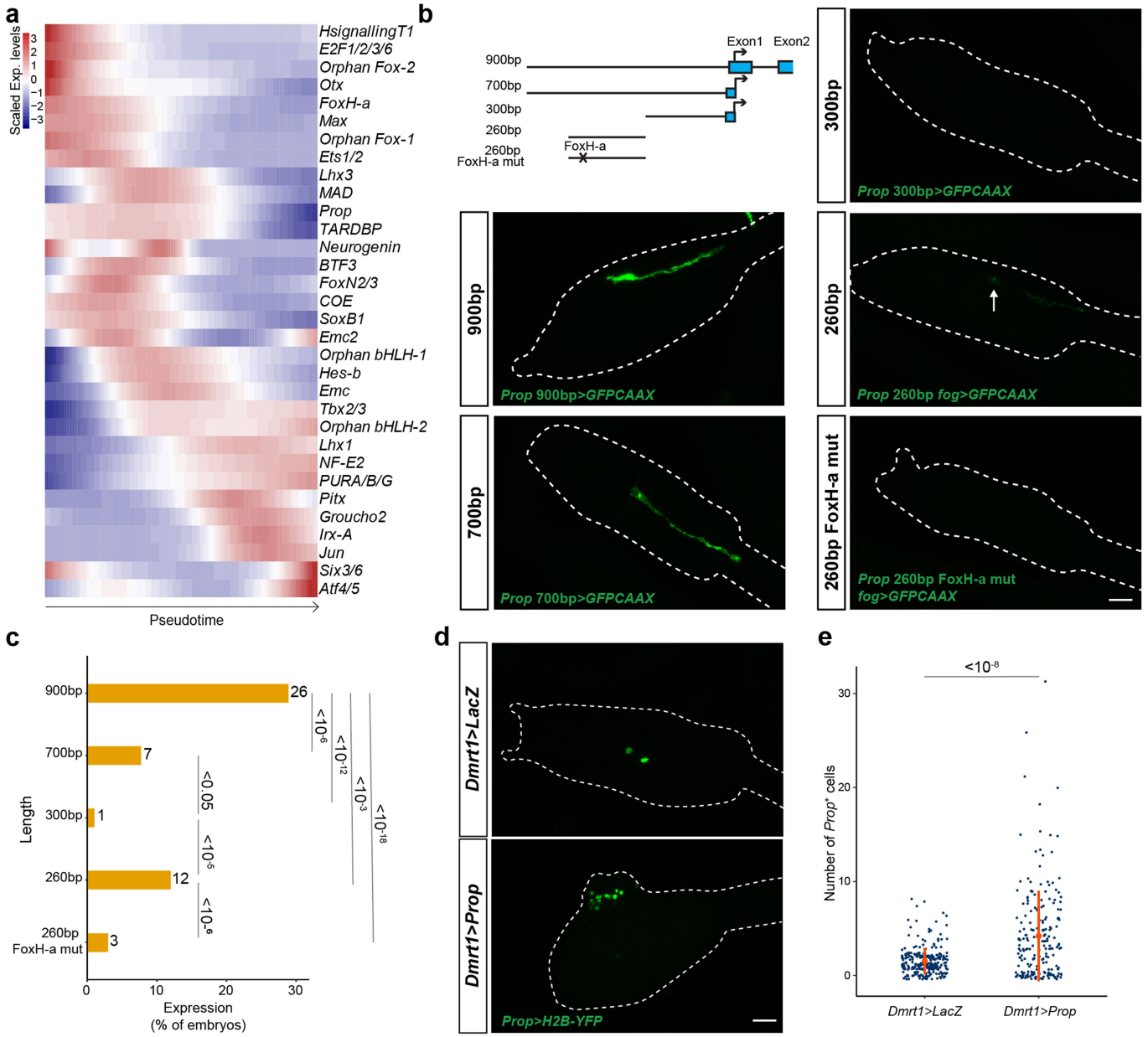


Extended Data Fig. 7 | See next page for caption.

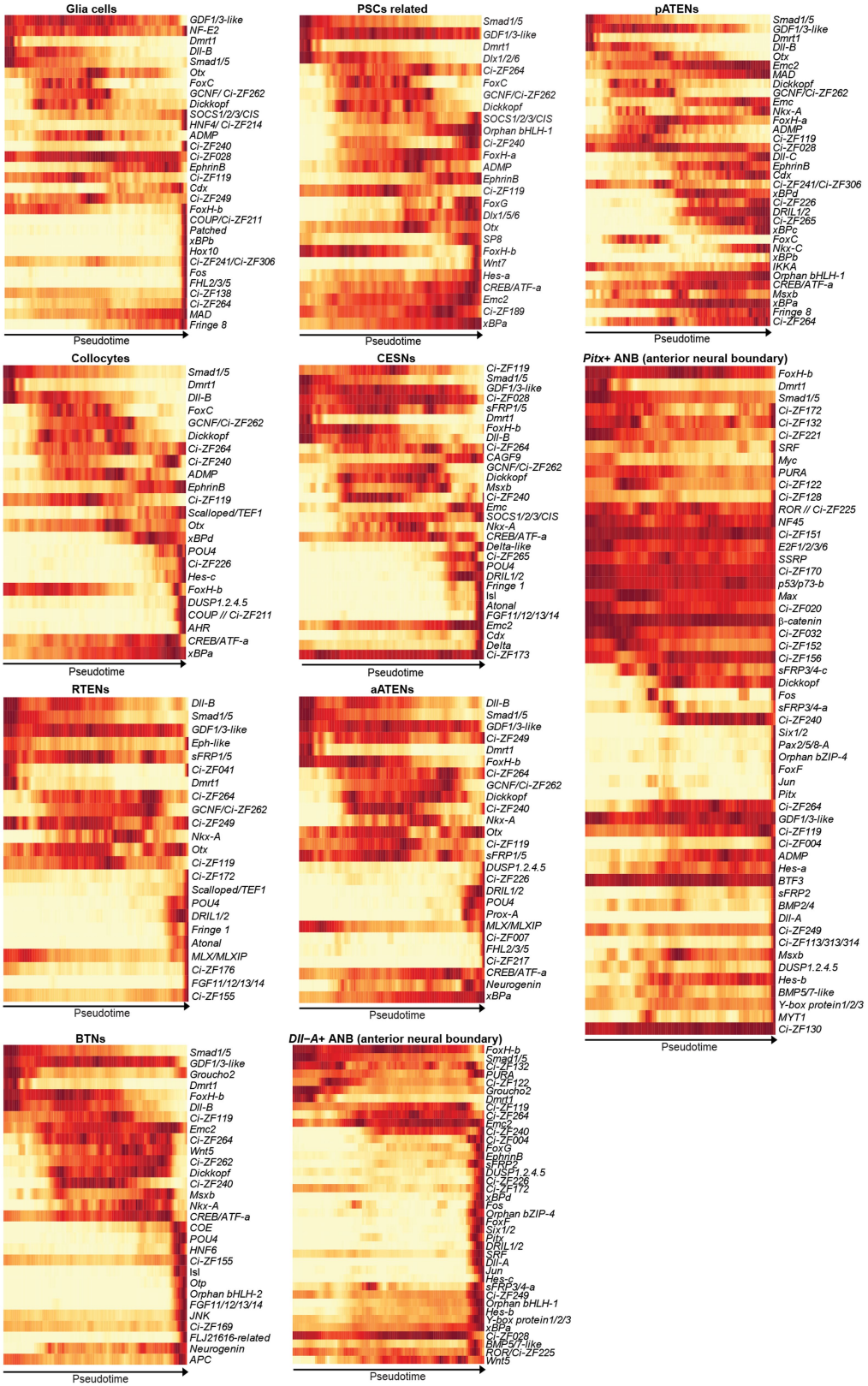
Extended Data Fig. 7 | The identification of rare neuronal subtypes in the larval stage. a, Distribution of cells that express *Galr2* in the *t*-SNE plot. Cells within the dashed circle show *Galr2* expression in bipolar tail neurons ($n = 26$ cells). Reporter assay with a bipolar-tail-neuron minimal enhancer for *Galr2* shows the specific activity of *Galr2* in bipolar tail neurons ($n = 3$ electroporation experiments). **b**, Distribution of cells that express *Dmbx* in the *t*-SNE plot. Cells within the dashed circle show *Dmbx* expression in decussating neurons ($n = 4$ cells). The 5′ regulatory sequences of *Dmbx* are active in decussating neurons (red, $n = 3$ electroporation experiments). **c**, Distribution of cells that express *NP* in the *t*-SNE plot. Cells within the dashed circle show *NP* expression in VP^+ posterior sensory vesicle ($n = 11$ cells). Reporter assay for *NP* (green) shows the specific expression of *NP* in neurons in the posterior sensory vesicle ($n = 3$ electroporation experiments). **d**, Distribution of cells that express *Prop* in the *t*-SNE plot. Cells within the dashed circle show *Prop* expression in Eminens neurons ($n = 17$ cells). Expression of the *Prop* reporter gene is specific to Eminens neurons (green) ($n = 3$ electroporation experiments). **e**, *t*-SNE plot of the larval nervous system showing cells that express *Ptf1a* (top) and *VP* (bottom). The dotted circle corresponds to coronet cells (top, $n = 72$ cells) and VP^+ posterior sensory vesicle cluster (bottom, $n = 11$ cells). **f**, Expression of the reporter *Ptf1a>mChCAAX* (red) for coronet cells and *NP>GFPCAAX* (green) for VP^+ posterior sensory vesicle shows that these cell populations do not contact each other, but are in close vicinity (top, $n = 3$ electroporation experiments; the GFP channel is shown in **c**). Expression of the reporter *Prop>mChCAAX* (red) for Eminens neurons and *NP>GFPCAAX* (green) for VP^+ posterior sensory vesicle. NP^+ cells are also in proximity to Eminens neurons (bottom, $n = 2$ electroporation experiments). Scale bars, 10 μm.



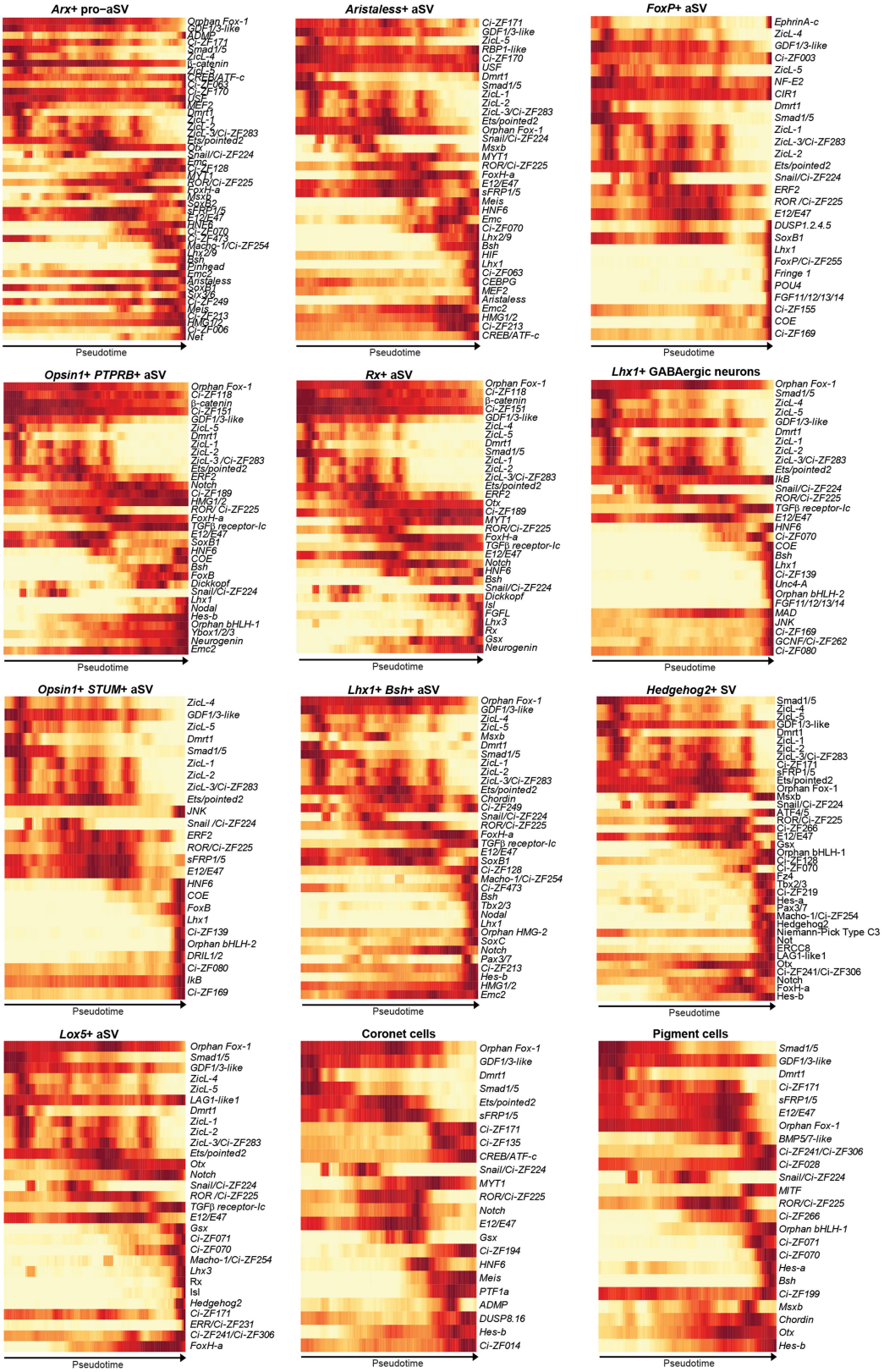
Extended Data Fig. 8 | Expression of marker genes for Eminens neurons. a–h, Expression levels of eight marker genes in the larval nervous system, shown in *t*-SNE plots (left, $n = 1{,}704$ cells), and their corresponding reporter assays (*mChCAAX* for *vGat* and *H2B-mCherry* for the other genes, red) with a *Prop>GFPCAAX* reporter (green, right). $n = 2$ electroporation experiments for *Gad*, *S39aa* 2.2 kb, *Znt3* and *Asic1b*; $n = 3$ electroporation experiments for *vGat*, *Calm*, *Fgf13* and *Galr2*. The dashed circle in the *t*-SNE plots identifies Eminens neurons. Scale bars, 20 μm.



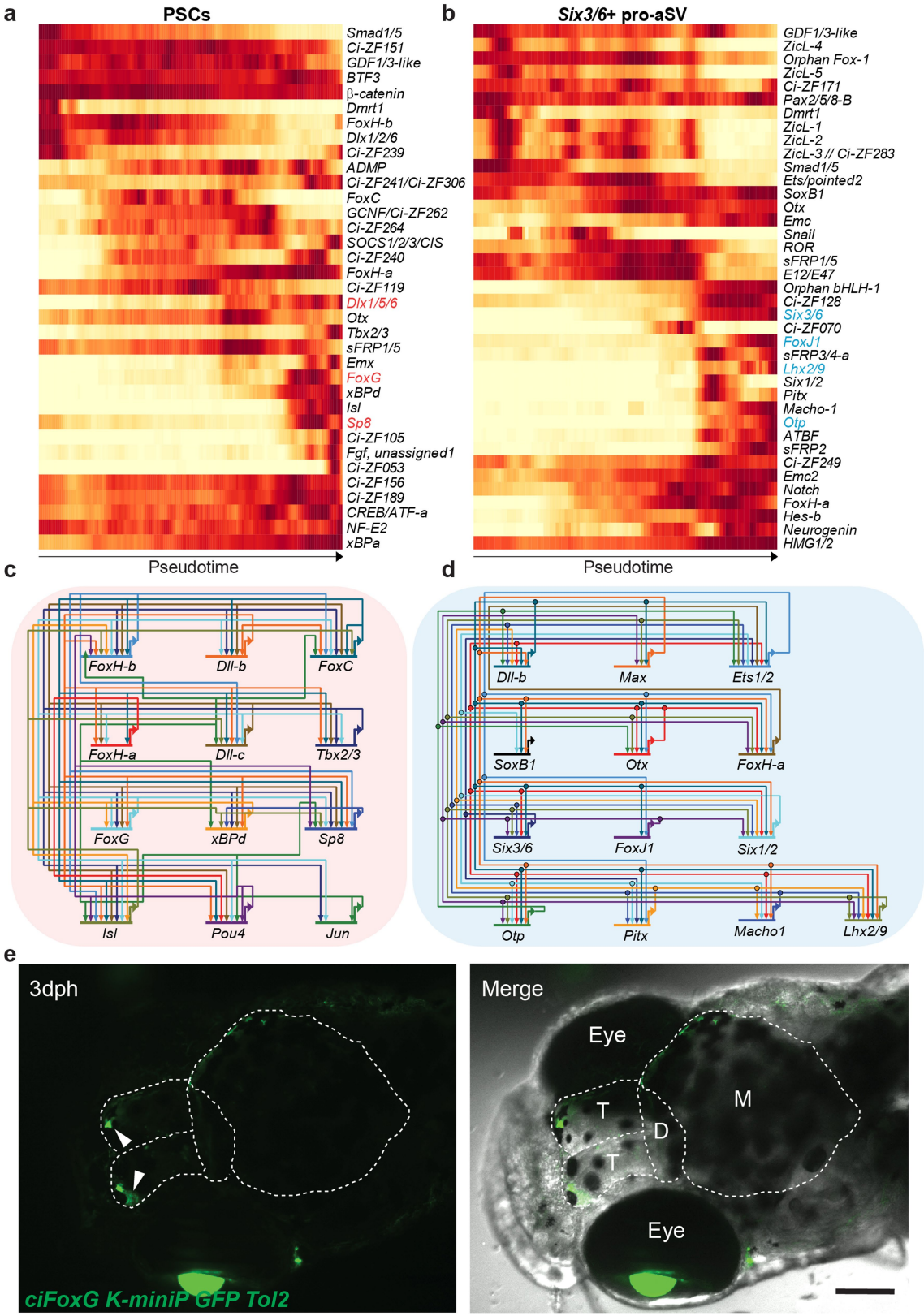
Extended Data Fig. 9 | Manipulation of Eminens gene regulatory network. a, Pseudotemporal expression profiles of regulatory genes and signalling components in Eminens neurons. **b**, Diagram of the *Prop* regulatory sequences with their length indicated on the left. A representative embryo is shown for the different fusion genes (GFPCAAX, green). The minimal *Prop* enhancer has weak expression in Eminens neurons (arrow). When the binding site for FoxH-a was mutated (260 bp FoxH-a mut), these regulatory sequences show even less activity. **c**, Bar plot of the percentage of the embryos that express GFP shown in **b**. Numbers on the right of the column correspond to the percentage of GFP$^+$ embryos. χ^2 test with four degrees of freedom was performed ($P < 2.2 \times 10^{-16}$), followed by two-sided Fisher's exact test with Bonferroni adjustment for multiple comparisons. *P* values: 900 bp versus 700 bp, $P = 1.05 \times 10^{-7}$; 900 bp versus 300 bp, $P = 3.47 \times 10^{-13}$; 900 bp versus 260 bp, $P = 2.36 \times 10^{-4}$; 900 bp versus 260-bp FoxH-a mut, $P = 1.81 \times 10^{-19}$; 700 bp versus 300 bp, $P = 0.011$; 700 bp versus 260 bp, $P = 0.36$; 700 bp versus 260-bp FoxH-a mut, $P = 0.088$; 300 bp versus 260 bp, $P = 5.59 \times 10^{-6}$; 300 bp versus 260-bp FoxH-a mut, $P = 0.69$; 260 bp versus 260-bp FoxH-a mut, $P = 1.27 \times 10^{-7}$. Numbers of embryos: 900 bp, $n = 207$; 700 bp, $n = 300$; 300 bp, $n = 160$, all pooled over 2 electroporation experiments; 260 bp, $n = 440$, 260-bp FoxH-a mut, $n = 750$, all pooled over 3 electroporation experiments. **d**, Overexpression of *Prop* using *Dmrt1* regulatory sequences causes supernumerary *Prop*$^+$ cells (bottom panel) compared to control embryos expressing *LacZ* (top). The 2-kb *Prop* reporter gene shows specific expression in Eminens neurons (H2B–YFP, green). The images show representative embryos for both conditions. **e**, Quantification of *Prop*$^+$ cells from the experiments in **d**. *Dmrt1>LacZ*, $n = 269$ embryos; *Dmrt1>Prop*, $n = 210$ embryos, pooled over 3 electroporation experiments. The orange dots indicate the mean and the bars indicate the s.d. *Dmrt1>LacZ*, 1.5 ± 1.4 cells; *Dmrt1>Prop*, 4.2 ± 4.8 cells. Mann–Whitney *U*-test, $P = 3.65 \times 10^{-9}$. Scale bars, 20 μm.



Extended Data Fig. 10 | Pseudotemporal gene-expression cascade of the peripheral nervous system. Representative transcription factors and signalling pathway genes along pseudotime in the reconstructed developmental trajectories of the peripheral nervous system are shown.



Extended Data Fig. 11 | Pseudotemporal gene-expression cascade of the central nervous system of a-lineage. Representative transcription factors and signalling pathway genes along pseudotime in reconstructed developmental trajectories of the central nervous system of a-lineage are shown.



Extended Data Fig. 12 | Model for the evolution of the telencephalon. **a**, Gene-expression cascade of regulatory genes and signalling components of palp sensory cells (also known as axial columnar cells). Genes implicated in the development of the vertebrate telencephalon are labelled in red. **b**, Gene-expression cascade of regulatory genes and signalling components in the anterior-most regions of the sensory vesicle (*Six3/6*$^{+}$ pro-anterior sensory vesicle). Genes implicated in vertebrate telencephalon development are labelled in blue. **c**, **d**, The putative regulatory interactions among transcription factors from the cascade of palp sensory cells (**c**) and *Six3/6*$^{+}$ pro-anterior sensory vesicle (**d**) along their developmental trajectories. **e**, The *FoxG* reporter gene with *Ciona* enhancer sequence exhibits restricted expression in a subset of cells in the olfactory bulb of the killifish telencephalon (arrowheads) and in the eye lens (left, GFP channel; right, merged image of bright-field and GFP channel images). $n = 3$ independent transgenic lines (Methods). D, diencephalon; M, midbrain; T, telencephalon. Scale bar, 400 μm.

nature research

Corresponding author(s): Michael Levine, Kai Chen

Last updated by author(s): May 23, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Raw sequencing reads were filtered by Illumina HiSeq Control Software and only pass-filter reads were used for further analysis. Samples were run on both lanes of a HiSeq 2500 Rapid Run mode flow cell. Base calling was performed by Illumina RTA version 1.18.64.0. BCL files were then converted to FASTQ format using bcl2fastq version 1.8.4 (Illumina). Reads that aligned to phix (using Bowtie version 1.1.1) were removed as well as reads that failed Illumina's default chastity filter. We then combined the FASTQ files from each lane and separated the samples using the barcode sequences allowing 1 mismatch (using barcode_splitter version 0.18.2). Using 10x CellRanger version 2.0.1, the count pipeline was run with default settings on the FASTQ files to generate gene–barcode matrices for each sample.
Data analysis	For dimensional reduction, clustering and t-SNE visualization, Seurat v2.3.4 was applied with an implement of a modified Fast Fourier Transform-accelerated Interpolation-based t-SNE method. In order to capture the developmental transitions stemming from different blastomeres at 110 cell stage, we performed "ancestor voting" between clusters across time as described in Briggs, J. A. et al. 2018. For notochord and Eminens cells, we employed monocle 2 to construct the single cell trajectory. For tissues that harbored more complexity during development, such as the mesenchyme and nervous system, we employed a simulated diffusion-based computational reconstruction method, URD, for acquiring the transcriptional trajectories during embryogenesis. Cluster-buster was used to find clusters of pre-specified motifs in 2kbp upstream of the TSS of each gene.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Raw sequencing data and gene expression matrix are available in Gene Expression Omnibus (GEO) under accession number GSE131155

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the sequencing, 100 to 500 embryos per samples were collected for following dissociation and scRNA-Seq. For the functional analysis and reporter assay, 18 to 610 embryos per sample were used. For the killifish reporter assay, between 25 and 46 embryos per lines were analyzed.
Data exclusions	To remove signals from putative empty droplet or degraded RNA, low-quality transcriptomes were filtered for each time course sample, as follows: 1) we discarded cells with less than 1000 expressed genes; 2) Cells with UMIs exhibiting five SDs above the mean were not included in our analyses (Supplementary Table 1); 3) we only consider genes that were expressed in at least 3 cells in each dataset. In total, 90,579 cells were kept for the following analysis.
Replication	We performed 2 biological replicates for stages from initial gastula to late tail bud II, and 3 replicates for swimming tadpole stage. All the reporter assays were performed at least twice. Two replicates were done for the minimal Prop enhancer assay. Three replicates were done for the overexpression assay as well as for the mutation assay on Prop minimal enhancer. Three different lines of transgenic killifish were generated for the reporter assay.
Randomization	Embryos in experiment were randomized and collected for dissociation before cell indexing on 10X Genomics Chromium system.
Blinding	Investigators were blinded to group allocation during data collection and analysis: embryo collection and scRNA-Seq data analysis were performed by two different researchers. For functional assays, no particular blinding strategy was adopted. Experiments were performed by one researcher.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Experiment of African killifish N. furzeri were performed using the GRZ strain
Wild animals	Ciona intestinalis were purchased from M-REP, San Diego, California, which collected them in San Diego area.

Field-collected samples	OR
Ethics oversight	Ciona intesitnalis are non-vertebrates; Work with killifish was performed according to guidelines of the Stowers Institute for Medical Research.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ARTICLE

https://doi.org/10.1038/s41586-019-1367-0

Somatic mutations and cell identity linked by Genotyping of Transcriptomes

Anna S. Nam[1,2,3,16], Kyu-Tae Kim[2,3,4,16], Ronan Chaligne[2,3,4,16], Franco Izzo[2,3,4], Chelston Ang[2,3,4], Justin Taylor[5], Robert M. Myers[2,3,4,6], Ghaith Abu-Zeinah[4,7], Ryan Brand[2,3,4], Nathaniel D. Omans[2,3,4,8], Alicia Alonso[9], Caroline Sheridan[9], Marisa Mariani[9], Xiaoguang Dai[10], Eoghan Harrington[10], Alessandro Pastore[5], Juan R. Cubillos-Ruiz[11], Wayne Tam[1], Ronald Hoffman[12], Raul Rabadan[13], Joseph M. Scandura[3,7], Omar Abdel-Wahab[5], Peter Smibert[14,17] & Dan A. Landau[2,3,4,15,17]*

Defining the transcriptomic identity of malignant cells is challenging in the absence of surface markers that distinguish cancer clones from one another, or from admixed non-neoplastic cells. To address this challenge, here we developed Genotyping of Transcriptomes (GoT), a method to integrate genotyping with high-throughput droplet-based single-cell RNA sequencing. We apply GoT to profile 38,290 CD34$^+$ cells from patients with *CALR*-mutated myeloproliferative neoplasms to study how somatic mutations corrupt the complex process of human haematopoiesis. High-resolution mapping of malignant versus normal haematopoietic progenitors revealed an increasing fitness advantage with myeloid differentiation of cells with mutated *CALR*. We identified the unfolded protein response as a predominant outcome of *CALR* mutations, with a considerable dependency on cell identity, as well as upregulation of the NF-κB pathway specifically in uncommitted stem cells. We further extended the GoT toolkit to genotype multiple targets and loci that are distant from transcript ends. Together, these findings reveal that the transcriptional output of somatic mutations in myeloproliferative neoplasms is dependent on the native cell identity.

Somatic mutations underlie the development of clonal outgrowth, malignant transformation[1] and subclonal diversification[2–5]. Nonetheless, clonally derived populations often lack cell-surface markers that distinguish them from normal cells or that can help to distinguish subclones, which limits our ability to link the clonal architecture of malignant populations with transcriptional read-outs. For example, although myeloproliferative neoplasms result from recurrent somatic mutations in *CALR*, *JAK2* and *MPL*[6,7], the mutated clone often represents a subset of bone marrow progenitors that lack distinctive surface markers to distinguish them from non-neoplastic progenitors. Thus, we are unable to study the effect of myeloproliferative neoplasm mutations in the context of the haematopoietic progenitor subtype identity.

Although advanced methods have been developed to capture both transcriptional information and genotype at the single-cell level[8,9], these methods often lack the throughput required to study complex systems such as haematopoietic differentiation. Droplet-based sequencing enables the transcriptomic profiling of thousands of cells[10,11], and can potentially also provide genotypic information regarding coding mutations (Extended Data Fig. 1a). However, current methods—by design—provide sequence information for only a short fragment at the transcript end, which limits the ability of these techniques to jointly genotype somatic mutations. To overcome this challenge, we developed GoT to link the genotyping of expressed genes to transcriptional profiling of thousands of single cells. We applied GoT to CD34$^+$ cells from patients with myeloproliferative neoplasms, which revealed that myeloproliferative neoplasm mutations in haematopoietic progenitor cells do not lead to uniform transcriptional outputs—but instead show a strong dependence on the progenitor cell identity.

Somatic genotyping in droplet scRNA-seq

To link genotypes to single-cell RNA sequencing (scRNA-seq) in high-throughput droplet-based platforms, we modified the 10x Genomics platform to amplify the targeted transcript and locus of interest (Fig. 1a, Extended Data Fig. 1b, c, Methods). We then investigated amplicon reads for mutational status, and linked the genotype to single-cell gene-expression profiles using shared cell barcodes (Extended Data Fig. 2a, b). We tested the ability of GoT to co-map single-cell genotypes and transcriptomes in a species-mixing experiment, in which mouse cells with a mutant human *CALR* transgene were mixed with human cells with a wild-type human *CALR* transgene[12] (Fig. 1b). Consistent with precision genotyping, the vast majority of cells with transcripts aligned to the mouse genome showed mutant *CALR* whereas cells with transcripts aligned to the human genome showed wild-type *CALR* (96.7% of cells matched the expected species) (Fig. 1b, Extended Data Fig. 2c–g).

CALR mutations have previously been demonstrated to activate MPL (which results in megakaryocytic proliferation)[7,12–16], but how the mutations perturb the early differentiation of haematopoietic stem and progenitor cells (HSPCs) is largely unknown. We therefore applied GoT to CD34$^+$ bone marrow cells from five patients with *CALR*-mutated essential thrombocythaemia, who had not been treated

[1]Department of Pathology and Laboratory Medicine, Weill Cornell Medicine, New York, NY, USA. [2]New York Genome Center, New York, NY, USA. [3]Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. [4]Division of Hematology and Medical Oncology, Department of Medicine, Weill Cornell Medicine, New York, NY, USA. [5]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [6]Tri-Institutional MD-PhD Program, Weill Cornell Medicine, Rockefeller University, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [7]Richard T. Silver MD Myeloproliferative Neoplasms Center, Division of Hematology and Medical Oncology, Department of Medicine, Weill Cornell Medicine, New York, NY, USA. [8]Tri-Institutional Training Program in Computational Biology and Medicine, Memorial Sloan Kettering Cancer Center, Cornell University, Weill Cornell Medicine, New York, NY, USA. [9]Epigenomics Core Facility, Weill Cornell Medicine, New York, NY, USA. [10]Oxford Nanopore Technologies, New York, NY, USA. [11]Department of Obstetrics and Gynecology, Weill Cornell Medicine, New York, NY, USA. [12]Division of Hematology and Medical Oncology, Department of Medicine, Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA. [13]Department of Systems Biology, Columbia University Medical Center, New York, NY, USA. [14]Technology Innovation Lab, New York Genome Center, New York, NY, USA. [15]Institute for Computational Biomedicine, Weill Cornell Medicine, New York, NY, USA. [16]These authors contributed equally: Anna S. Nam, Kyu-Tae Kim, Ronan Chaligne. [17]These authors jointly supervised this work: Peter Smibert, Dan A. Landau. *e-mail: dlandau@nygenome.org

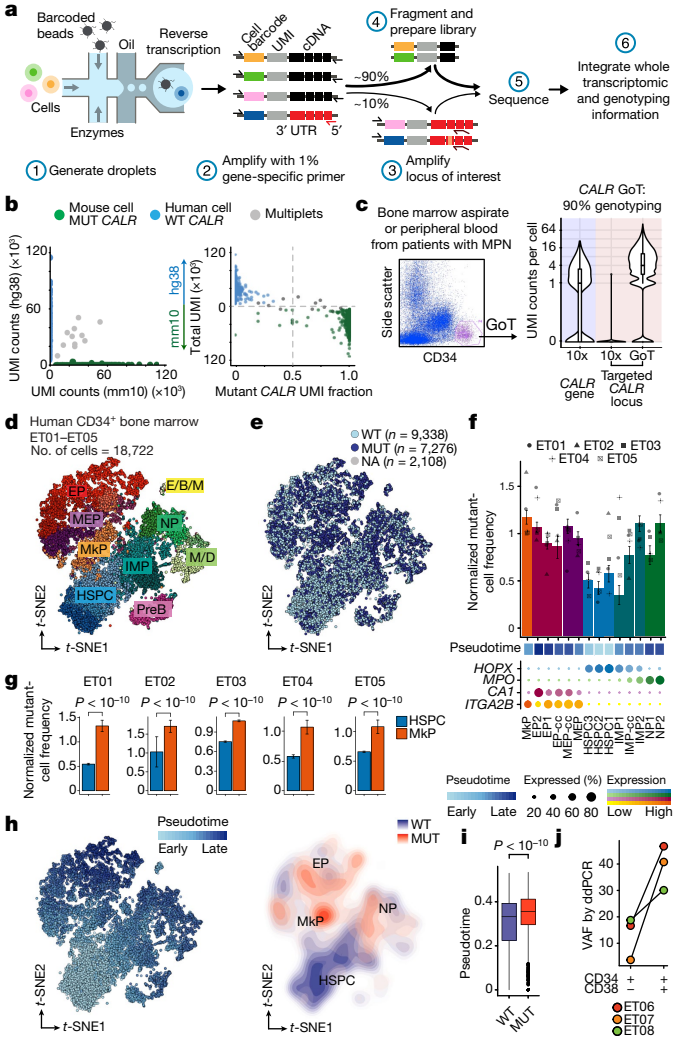


Fig. 1 | GoT provides genotyping of somatic mutations for thousands of cancer cells, and reveals a differential fitness effect of *CALR* mutation in subsets of haematopoietic progenitor cells. a, Schematic of GoT workflow. UMI, unique molecular identifier; UTR, untranslated region. **b**, Species-mixing study with mutant (MUT) *CALR* mouse cells and wild-type (WT) *CALR* human cells. Left, 10x reads from singlet cells map to human or mouse genome. Right, mouse versus human genome alignment of 10x data (*y* axis) and GoT data (*x* axis; $n = 1{,}259$ cells). **c**, FACS of $CD34^+$ cells (left) and UMIs per cell (right) for *CALR* transcript (blue shade) or targeted locus (pink shade) from a representative sample ET01 ($n = 6{,}811$ cells). See Extended Data Fig. 3a, b for similar results across samples ET01–ET05 and MF01–MF05. MPN, myeloproliferative neoplasms. **d**, **e**, *t*-distributed stochastic neighbour embedding (*t*-SNE) projection of $CD34^+$ cells from patients with essential thrombocythaemia with cluster assignment (**d**) and genotyping data (**e**). E/B/M, eosinophil, basophil and mast cell progenitors; EP, erythroid progenitors; IMP, immature myeloid progenitors; M/D, monocyte–dendritic cell progenitors; MEP, megakaryocytic–erythroid progenitors; MkP, megakaryocytic progenitors; NP, neutrophil progenitors; PreB, precursor B cells; NA, not assignable. **f**, Normalized frequency of mutant cells (Methods). Bars show aggregate analysis of samples ET01–ET05 with mean ± s.d. of 100 downsampling iterations to 1 genotyping UMI per cell. Points represent mean of $n = 100$ downsampling iterations for each sample. cc, cell cycle. **g**, Normalized frequency of mutant cells. Mean ± s.d. of $n = 100$ downsampling iterations (Wilcoxon rank-sum test, two-sided). **h**, *t*-SNE projection of essential thrombocythaemia $CD34^+$ cells with pseudotime (left) and density plot of wild-type and mutant cells (right). **i**, Pseudotime in wild-type versus mutant cells. *P* value from likelihood ratio test of linear mixed model with or without mutation status (Methods). **j**, Bulk VAF of *CALR* mutation in FACS-sorted cells from patients with essential thrombocythaemia by ddPCR. In all figures, box plots represent the median, bottom and top quartiles, whiskers correspond to 1.5 × the interquartile range; violin plots depict kernel density estimates to show the density distribution.

with disease-modifying therapy at the time of biopsy (Supplementary Table 1). *CALR* genotyping data were available for 16,614 of 18,722 cells (88.7%), compared to only 1.4% by investigation of *CALR* in the conventional 10x Genomics data (Fig. 1c, Extended Data Fig. 3a–d).

To investigate the cellular identities of these progenitors, we performed clustering that was agnostic to the genotyping information, based on the transcriptome information alone[17,18] (Fig. 1d, Extended Data Fig. 4a–c). The projection of genotypes onto progenitor maps demonstrated that mutated cells involved all $CD34^+$ stem and progenitor clusters, which is consistent with previous[6] bulk PCR analysis of *CALR* in $CD34^+$ cell subsets sorted by fluorescence-activated cell sorting (FACS) (Fig. 1e; see Extended Data Fig. 4d, e for validation with an alternative clustering framework[19]). Notably, mutated cells did not form novel independent clusters, which confirms that scRNA-seq alone cannot distinguish mutant from wild-type cells and demonstrates that *CALR* mutations in essential thrombocythaemia affect the entire haematopoietic differentiation hierarchy.

Progenitor-specific *CALR* mutant fitness

Although mutant cells were observed across all progenitor clusters, their frequencies varied between clusters. The frequency of *CALR*-mutated cells was higher in committed myeloid progenitors (Fig. 1f) and especially in megakaryocytic progenitors (MkPs), which are closely associated with the disease phenotype of elevated platelet counts, as compared to uncommitted HSPC clusters across samples ($P < 10^{-10}$, linear mixed model) (Fig. 1g, Extended Data Fig. 5a). Consistently, pseudotemporal ordering (pseudotime) analysis[20,21] showed that *CALR*-mutated cells were enriched in cells at later pseudotime points, compared to wild-type cells ($P < 10^{-10}$, linear mixed model) (Fig. 1h, i, Extended Data Fig. 5b, c). We orthogonally validated this finding using bulk genomic DNA droplet-based digital polymerase chain reaction (ddPCR), which showed that there was a lower *CALR* variant allele frequency (VAF) in $CD34^+CD38^-$ HSPCs compared to $CD34^+CD38^+$ progenitors ($P = 0.02$, Wilcoxon rank-sum test) (Fig. 1j). Thus, although *CALR* mutations arise in uncommitted haematopoietic stem cells (HSCs) and therefore propagate to populate the entire differentiation tree, the effect of *CALR* mutations on fitness increases with myeloid differentiation (Extended Data Fig. 5d, e).

GoT enables direct comparison of transcriptional programs between mutant and wild-type cells—not only within the same sample, but also within the same progenitor cluster. For example, *CALR*-mutated progenitors displayed increased expression of genes that are upregulated in ex vivo cultured progenitors from *JAK2*-mutated essential thrombocythaemia[22] (most significantly in MkP clusters; combined $P < 10^{-10}$, Fisher's method) (Fig. 2a, b), as *JAK2* and *CALR* mutations partially converge through the activation of similar downstream pathways[23]. We therefore reasoned that a progenitor-subtype-specific comparison of cell-cycle gene expression[24] (Supplementary Table 2) in mutant versus wild-type cells might reveal whether the lower fitness effect of *CALR* mutations in HSPCs compared with MkPs stems from differences in cell proliferation. Whereas mutant HSPCs exhibited only a modest increase in cell-cycle gene expression compared to wild type ($P = 0.015$, Wilcoxon rank-sum test, mean fold change of 1.2, 95% confidence interval, 1.1–1.4), mutant MkPs demonstrated a robust increase in cell-cycle gene expression versus their wild-type counterparts ($P = 4.4 \times 10^{-4}$, Wilcoxon rank-sum test, fold change 1.8, 95% confidence interval 1.4–2.8) (Fig. 2c, Extended Data Fig. 6a). Notably, the degree of increase in cell-cycle gene expression in mutant MkPs correlated with the platelet counts of the patients (Fig. 2d), which suggests that early progenitor cell characteristics may correlate with clinical phenotypes and has the potential to inform our understanding of patient-to-patient variability that occurs in spite of shared mutated genotypes.

Cell-to-cell variation exists even within progenitor clusters. For example, MkPs represent a heterogeneous population that is composed of less-differentiated cells with higher expression of HSC genes[25] ($HSC^{high}MkP^{low}$), and more-committed MkP cells that display

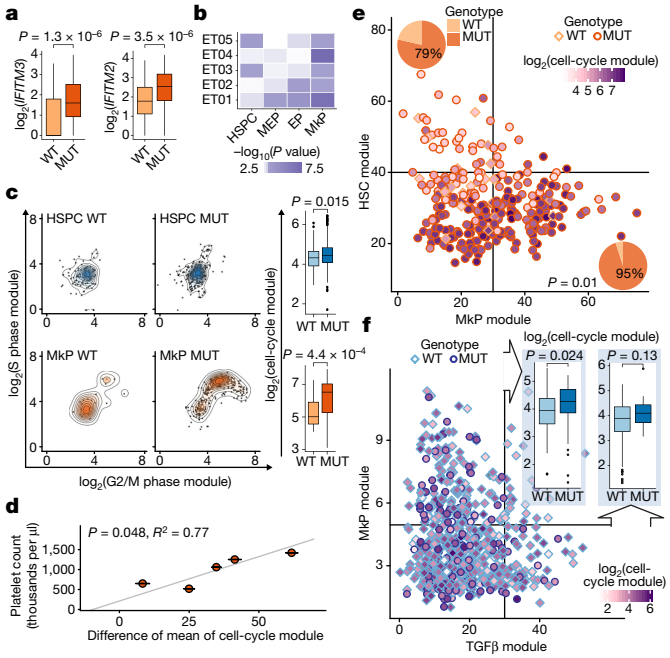


Fig. 2 | *CALR* mutations result in a higher proliferative effect on MkPs compared to HSPCs. a, Expression of representative genes upregulated in *JAK2*-mutated essential thrombocythaemia cultured cells[22] in *CALR* wild-type ($n = 157$) versus mutant ($n = 85$) MkPs from a representative sample ET01. **b**, Heat map of $-\log_{10}(P$ value) from comparisons (between mutant and wild-type cells) of genes expressed in *JAK2*-mutated essential thrombocythaemia cells (Supplementary Table 6). **c**, Cell-cycle module expression in HSPCs ($n = 108$ wild type versus $n = 240$ mutant) and MkPs ($n = 25$ wild type versus $n = 276$ mutant) from ET03 (Extended Data Fig. 6a). **d**, Platelet counts versus difference of mean cell-cycle score (± s.e.m.) between wild-type and mutant MkPs ($n = 5$ samples; *F*-test). **e**, Expression of MkP and HSC modules in MkPs from sample ET03. Pie charts of wild-type versus mutant cell frequencies in $HSC^{low}MkP^{high}$ ($n = 121$ cells) and $HSC^{high}MkP^{low}$ ($n = 28$ cells) populations. Fisher's exact test, two-sided. **f**, Expression of TGFβ and MkP modules in HSPCs from sample ET01, and cell-cycle score in HSPCs in $MkP^{high}TGF\beta^{low}$ ($n = 127$ wild type versus $n = 41$ mutant) and $MkP^{low}TGF\beta^{high}$ ($n = 105$ wild type versus $n = 15$ mutant) populations. *P* values for **a**–**c**, **f** are from a two-sided Wilcoxon rank-sum test.

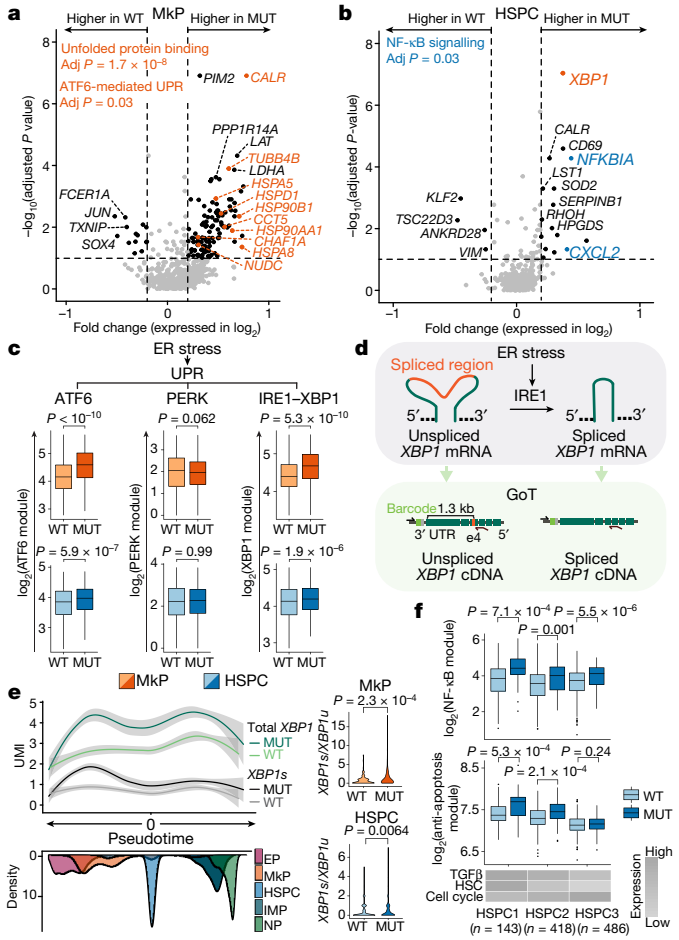


Fig. 3 | Transcriptional effects of *CALR* mutation are dependent on cell identity. a, **b**, Differentially expressed genes between mutant and wild-type MkPs (**a**) and between mutant and wild-type HSPCs (**b**) across ET01–ET05 samples (Supplementary Table 6). *P* values combined using Fisher's combined test with Benjamini–Hochberg adjustment. Key gene-set enrichments are shown (hypergeometic test, Methods). **c**, Expression of genes upregulated in UPR branches in MkPs ($n = 442$ wild type versus $n = 640$ mutant) and HSPCs ($n = 1{,}704$ wild type versus $n = 613$ mutant) from samples ET01–ET05. *P* values from likelihood ratio tests of linear mixed model, with and without mutation status (Methods). ER, endoplasmic reticulum. **d**, Schematic of GoT applied to the *XBP1* splice site. **e**, Left, local regression of total and spliced *XBP1* (*XBP1s*) expression in progenitor cells from samples ET03 and ET04 ($n = 1{,}308$ wild type versus $n = 1{,}514$ mutant; shading denotes 95% confidence interval). Right, ratio of *XBP1s* to unspliced *XBP1* (*XBP1u*) in MkPs ($n = 115$ wild type versus $n = 248$ mutant) and HSPCs ($n = 489$ wild type versus $n = 302$ mutant). **f**, Expression of NF-κB pathway and anti-apoptosis genes in HSPC1 ($n = 116$ wild type versus $n = 27$ mutant), HSPC2 ($n = 365$ wild type versus $n = 53$ mutant) and HSPC3 ($n = 381$ wild type versus $n = 105$ mutant) from sample ET01. *P* values for **e**, **f** are from a two-sided Wilcoxon rank-sum test.

high expression of MkP-related genes[26] ($HSC^{low}MkP^{high}$) (Fig. 2e, Supplementary Table 2). Even within the MkP cluster, $HSC^{low}MkP^{high}$ cells showed increased cell-cycle gene expression and a higher frequency of mutant cells, compared to $HSC^{high}MkP^{low}$ cells ($P = 0.01$, Fisher's exact test). Similarly, mutant cells within the platelet-primed HSPCs ($MkP^{high}TGF\beta^{low}$) had higher cell-cycle gene expression than wild-type $MkP^{high}TGF\beta^{low}$ HSPCs, whereas mutant and wild-type HSPCs in a more-quiescent state ($MkP^{low}TGF\beta^{high}$) showed no difference in cell-cycle gene expression (Fig. 2f). These findings further emphasize that the effect of *CALR* mutations is dependent on cell state, and imparts a greater proliferative advantage in more-differentiated cells. These data also reveal that *CALR* mutations skew differentiation towards myeloid progenitors—including megakaryocytic priming—early during haematopoiesis (Extended Data Fig. 6b).

Cell identity and the effects of *CALR* mutations

GoT data further offer an opportunity for de novo discovery of differentially expressed genes, by examining wild-type versus mutant cells within the same subset of progenitors. Crucially, the wild-type cells serve as an ideal comparison set, because they share all potential environmental and patient-specific variables with the mutated cells. We identified 198 genes that were differentially expressed between mutant and wild-type MkPs (false-discovery-rate adjusted $P < 0.1$) (Fig. 3a, Supplementary Table 3). Mutated MkPs upregulated *HSPA5*, which encodes BiP, a key player in protein quality control that modulates the activities of the three transmembrane transducers of the unfolded protein response (UPR): PERK, IRE1 and ATF6[27]. Consistently, *CALR*-mutant MkPs showed upregulation of UPR genes (adjusted $P = 1.7 \times 10^{-8}$) and ATF6-mediated activation of chaperone genes (adjusted $P = 0.03$) (Fig. 3a, Supplementary Table 4), which provides direct in vivo validation of previous in vitro studies that have shown an increased UPR in *CALR*-mutated cells[28,29]. The UPR in this context may signal endoplasmic-reticulum stress in response to misfolded proteins, as the chaperone activity of CALR may be compromised by the mutation[30,31]. Notably, among 20 differentially expressed genes in mutant HSPCs, *XBP1*—an important regulator of the UPR—was upregulated in mutant cells, which suggests that UPR activation by

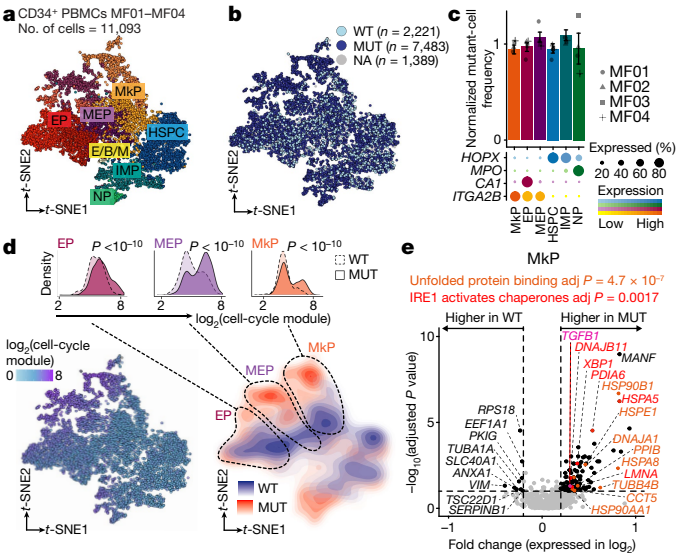


Fig. 4 | Effects of *CALR* mutation on haematopoietic progenitor cells from patients with myelofibrosis. a, **b**, *t*-SNE projection of CD34$^+$ cells from patients with myelofibrosis showing cluster assignment (**a**) and genotyping data from GoT (**b**). **c**, Normalized frequency of mutant cells (Methods). Bar graphs represent aggregate analysis of samples MF01–MF04, showing mean ± s.d. of 100 downsampling iterations to 1 genotyping UMI per cell. Grey points represent mean of 100 downsampling iterations for each sample. **d**, *t*-SNE projection of the CD34$^+$ cells, showing cell-cycle gene expression (left) and density plot of mutant and wild-type cells (right). Density plots of mutant versus wild-type cells along cell-cycle gene expression (inset, two-sided Wilcoxon rank-sum test) (Supplementary Table 6). **e**, Differentially expressed genes in mutant versus wild-type MkPs across samples MF01–MF04 (Supplementary Table 6). *P* values combined using Fisher's combined test with Benjamini–Hochberg adjustment. Key gene-set enrichments are shown (hypergeometic test, Methods).

CALR mutations extends to uncommitted progenitors (Fig. 3b). This is further supported by the upregulation of ATF6-target genes in mutant HSPCs, as compared to wild-type HSPCs[32] (Fig. 3c, Extended Data Fig. 7, Supplementary Table 2).

The UPR has previously been demonstrated to be preferentially mediated in HSPCs through PERK (which results in enhanced apoptosis upon endoplasmic-reticulum stress); this helps to eliminate endoplasmic-reticulum-stressed cells from the HSC pool[33]. By contrast, committed progenitors have robust activity of the IRE1–XBP1 arm of the UPR, which promotes survival through endoplasmic-reticulum-stress challenge[33]. We observed that in the specific context of endoplasmic-reticulum stress induced by *CALR* mutation, the PERK branch of the UPR was not enhanced[33] (Fig. 3c, Extended Data Fig. 7, Supplementary Table 2). By contrast, targets of XBP1 were upregulated in mutant MkPs ($P = 5.3 \times 10^{-10}$, linear mixed model) as well as in mutant HSPCs[34,35] ($P = 1.9 \times 10^{-6}$, linear mixed model) (Fig. 3c, Extended Data Fig. 7, Supplementary Table 2). As IRE1 catalyses the unconventional splicing of *XBP1* unspliced mRNA (*XBP1u*) into the active spliced form (*XBP1s*)[36] (Fig. 3d), we further validated that *CALR* mutations induce the activation of the IRE1 branch of the UPR, by repurposing GoT to probe for the spliced region of *XBP1* in single cells. *CALR* mutations robustly augmented the amount of *XBP1s* in MkPs (Fig. 3e). *CALR* mutations also resulted in an increased *XBP1s*/*XBP1u* ratio in HSPCs, which indicates IRE1 activity (Fig. 3e). These data thus suggest that, in the UPR induced by *CALR* mutations, IRE1 is activated in both HSPCs and MkPs, which skews stem and progenitor cells challenged by endoplasmic-reticulum stress towards survival.

Analysis of differential gene expression in HSPCs also revealed upregulation of the NF-κB pathway (adjusted $P = 0.03$), including upregulation of *CXCL2* and *NFKBIA* (Fig. 3b, Supplementary Tables 3, 4). Furthermore, upregulation of the NF-κB pathway gene set in mutant versus wild-type HSPCs was most notable in early uncommitted HSPCs (Fig. 3f, Supplementary Table 2). Mutant cells in this early HSPC subcluster also upregulated anti-apoptotic-related genes (Fig. 3f, Supplementary Table 2). As NF-κB pathway activation has previously been associated with anti-apoptotic effects[37] and HSC self-renewal[38], our data thus point to another potential mechanism that links *CALR* mutation and HSC outgrowth.

IRE1-mediated UPR in *CALR*-mutated myelofibrosis

As a proportion of patients with *CALR*-mutated essential thrombocythaemia eventually progress to myelofibrosis (or present initially with myelofibrosis), we examined whether *CALR* mutations impart a similar proliferative and survival advantage to progenitor cells from patients with *CALR*-mutated myelofibrosis, by examining 9,704 genotyped cells of 11,093 cells in total (87.5% genotyping rate) across four samples of myelofibrosis (Fig. 4a, b, Extended Data Fig. 8a). In contrast to essential thrombocythaemia, we did not observe enrichment of mutated cells in differentiated progenitors compared to HSPCs (Fig. 4c, Extended Data Fig. 8b), consistent with previous reports[39]. This suggests that, in the context of myelofibrosis, *CALR* mutations impart a strong fitness advantage even to HSPCs. Indeed, mutant cells were highly enriched in cell-cycle activity (Fig. 4d, Extended Data Fig. 8c).

As megakaryocytes have previously been demonstrated to have a principal role in the development of marrow fibrosis[40], we performed differential expression analysis between mutant and wild-type MkPs, which showed 92 differentially expressed genes (false-discovery-rate adjusted $P < 0.1$) (Fig. 4e, Supplementary Table 3). We identified the upregulation of *TGFB1*, which has previously been implicated in fibroblast stimulation by megakaryocytes[41,42], and thereby demonstrated that TGFβ production is dysregulated even in early progenitors. The upregulation of TGFβ signalling in mutant MkPs correlated with the degree of fibrosis in the bone marrow of the patient (Extended Data Fig. 8d). As in *CALR*-mutated essential thrombocythaemia, we observed a robust upregulation of UPR genes in mutated MkPs (adjusted $P = 4.7 \times 10^{-7}$) and, specifically, IRE1 activation (adjusted $P = 0.0017$) (Fig. 4e, Supplementary Table 4). Notably, a comparison between cycling and non-cycling wild-type cells did not show upregulation of the UPR, which affirms that activation of the UPR is not simply a byproduct of increased proliferation (Extended Data Fig. 8e, Supplementary Tables 3, 4). These findings suggest that the increased survival of *CALR*-mutant progenitors through the upregulation of the IRE1-mediated UPR is maintained through disease progression to myelofibrosis.

Multiplexed GoT for subclonal identities

Ongoing clonal evolution results in multi-clonal malignant populations, which require genotyping of multiple mutations in parallel. To test the ability of GoT to target multiple mutations, we targeted three mutations that affect *CALR* (VAF 43.5% by bulk exon sequencing), *NFE2* (VAF 33%) and *SF3B1* (VAF 47.5%) in 8,475 CD34$^+$ cells from a patient with myelofibrosis (Fig. 5a). The relative VAFs of these mutations suggest that this malignancy follows a nested (linear evolution) clonal structure, with a clonal *SF3B1* mutation and a progeny subclone containing a *CALR* mutation, which—in turn—has additional *NFE2*-mutated progeny (single-cell cloning validation in Extended Data Fig. 9a). GoT provided genotyping for *CALR* and *NFE2* in 74% and 60% of cells, respectively, and showed mutant frequencies (of 64% and 56%, respectively) comparable to those of single-cell cloning (85% and 71%, respectively, performed with unsorted peripheral blood cells).

In this context, GoT allows us to compare the transcriptional outputs of the different mutations alone, or in combination. For example, because *SF3B1* mutations have previously been shown to block erythroid maturation[43], we examined whether the addition of a *CALR* mutation would still confer increased proliferative status in megakaryocytic–erythroid progenitors. We found that the *SF3B1 CALR* double mutants exhibited an increased proliferative advantage over *SF3B1* single mutants (Fig. 5b), whereas the addition of *NFE2* mutation

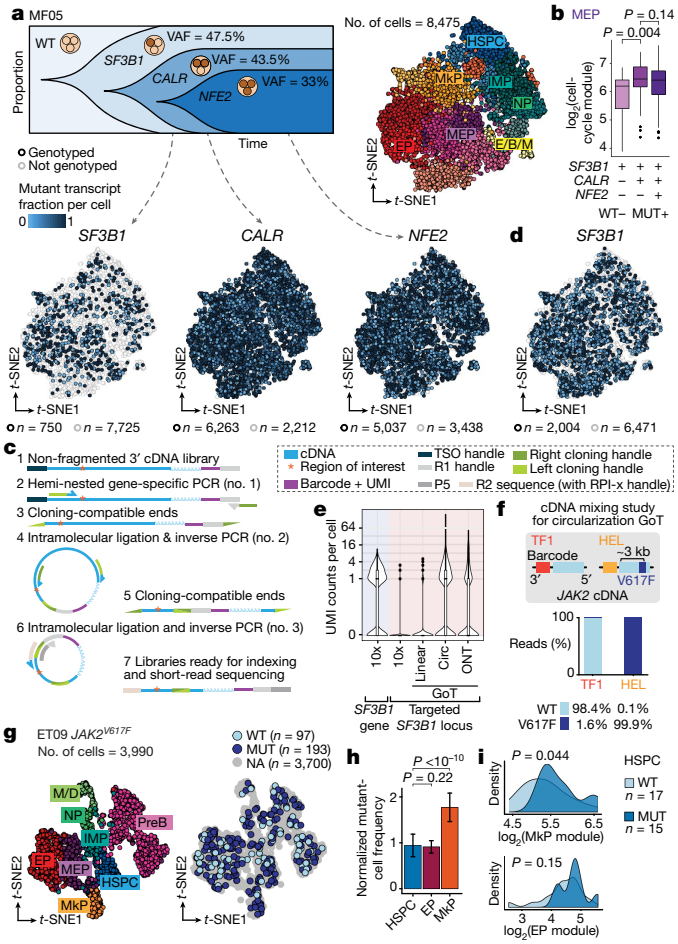


Fig. 5 | GoT dissects subclonal identity through multiplexing and targets loci that are distant from transcript ends via circularization. **a**, Schematic of clonal evolution of neoplastic cells from sample MF05 (top left). *t*-SNE projections of CD34$^+$ cells with cluster assignments (top right) and with GoT data for each variant (bottom). **b**, Cell-cycle score in subclonal megakaryocytic–erythroid progenitor populations ($n = 28$ single-mutant, 109 double-mutant and 293 triple-mutant cells). **c**, Schematic of circularization GoT. **d**, *t*-SNE projection of GoT data for *SF3B1* CD34$^+$ cells from sample MF05, from circularization GoT and linear GoT. **e**, UMIs per cell for *SF3B1* gene (blue shade) or targeted *SF3B1* locus (pink shade) from 10x, linear GoT sequenced on Illumina, circularization (circ) GoT and linear GoT sequenced with Oxford Nanopore Technology (ONT) ($n = 8{,}475$ cells). **f**, Mixing study with human *JAK2* wild-type cDNA from the TF-1 cell line and homozygous *JAK2*^{V617F} cDNA from the HEL cell line. Frequency of reads (wild type, V617F or not assignable) assigned to TF-1 or HEL cell barcodes. **g**, *t*-SNE projection of CD34$^+$ cells from a patient with *JAK2*^{V617F} essential thrombocythaemia, showing cluster assignment (left) and genotyping information (right) based on GoT data. **h**, Normalized frequency of mutant cells within the progenitor clusters (Methods). Mean ± s.d. of $n = 100$ downsampling iterations. **i**, Density plots of HSPCs along lineage-priming modules ($n = 17$ wild-type versus 15 mutant cells). *P* values for **b**, **h**, **i** are from a two-sided Wilcoxon rank-sum test.

(that is, the triple mutant) did not further increase cell-cycle activation. Thus, multiplexed GoT demonstrates the ability to examine complex clonal structures, as well as the need to assess the combinatorial transcriptional output of mutations in the context of the high-resolution mapping of cell identity.

Circularization GoT targets distant loci

GoT amplicon recovery is not only dependent on the expression level of the gene but also on the distance of the mutation locus from transcript ends (Extended Data Fig. 9b). Capture efficiency of a mutation that is distant from the 3′ end (>1.5 kb) (for example, *SF3B1* genotyping of 9% of cells, see Fig. 5a) was lower than for targets closer to the 3′ end. Although driver mutations are often found within 1.5 kb of one of the transcript ends (Extended Data Fig. 9c), loci of interest may reside at larger distances—and thus the dependency on relative proximity to transcript ends is limiting. We reasoned that the lower genotyping efficiency resulted at least in part from the inability of larger amplicon fragments to cluster efficiently on Illumina flow cells during sequencing. We further integrated our protocol with long-read sequencing using nanopore GridION X5 (Oxford Nanopore Technology), which demonstrated that *SF3B1* transcripts were captured accurately with our procedure (Extended Data Fig. 9d, e) and further confirmed low intra- and inter-transcript PCR recombination rate, even for these relatively large fragments (Extended Data Fig. 9f, g).

To overcome the limitation of amplicon fragment length, we applied sequential rounds of circularization and inverse PCR to remove the intervening sequence between the region of interest and the cell barcode, resulting in a fragment length compatible with short-read sequencing (Fig. 5c). Circularization GoT showed a high concordance with un-circularized GoT (that is, the standard linear GoT technique) for *CALR* genotyping (Extended Data Fig. 9h, i). When applied to the capture of *SF3B1* mutations, circularization GoT markedly increased the yield of genotyped cells from 750 to 2,004 cells (9% to 24% of cells) (Fig. 5d, e). These results demonstrate the ability of circularization GoT to extend our reach to targets that are distant from gene ends.

To further demonstrate the ability of circularization GoT to genotype efficiently even when mutations are at a considerable distance from a transcript end, we targeted *JAK2*^{V617F} (which is located about 2.3 kb from the closer transcript end). We first validated circularization GoT for *JAK2* via a mixing experiment using barcoded cDNA from the TF1 cell line (wild-type *JAK2*) and from HEL cells (homozygous *JAK2*^{V617F}), which showed accurate genotype assignment (Fig. 5f). Next, we genotyped primary CD34$^+$ cells from an individual with *JAK2*^{V617F} essential thrombocythaemia, and obtained genotyping information for 7.3% of cells (Fig. 5g) even for this gene, which is expressed at very low levels. Mutant-cell frequency was higher in MkPs than in HSPCs, whereas the mutant-cell frequency remained low in erythroid progenitors (Fig. 5h). This is concordant with the clinical phenotype of essential thrombocythaemia rather than polycythaemia vera (a disease that is associated with the same *JAK2* mutation, but which is characterized by erythrocytosis as the leading abnormality). Consistent with this observation, we observed a trend towards increased MkP priming in mutant HSPCs ($P = 0.04$, Wilcoxon rank-sum) (Fig. 5i, Supplementary Table 2)—albeit in a small number of genotyped HSPCs. These data suggest a skewing of differentiation towards megakaryopoiesis in HSPCs and may provide insights into the isolated megakaryocytic proliferation in *JAK2*-mutated essential thrombocythaemia.

Discussion

Here we present GoT, which captures both somatic genotypes and transcriptomic identities in thousands of single cells from primary cancer specimens. Building on previous experience with targeted amplification in droplet-based scRNA-seq[44,45], GoT overcomes the unique set of challenges presented by the genotyping of somatic mutations, including lower expression levels and large distances from the end of the sequenced transcripts.

GoT allowed us to directly investigate the transcriptional effect of *CALR* mutations in primary samples of myeloproliferative neoplasm, in which wild-type cells in the sample provide an ideal comparison set that controls for patient-specific and technical confounders. In essential thrombocythaemia, we observed that mutant *CALR* provided a greater fitness advantage through differentiation, which was associated with higher proliferation in committed myeloid progenitors than in uncommitted HSPCs. The ability of GoT to finely map transcriptional differences between mutant and wild-type HSPCs revealed an upregulation of NF-κB pathway genes in the most-undifferentiated mutant HSPCs, which supports a cell-intrinsic role for *CALR* mutation in NF-κB activation[46]. We further applied GoT to target the unconventional splice

site of *XBP1* to demonstrate that IRE1 activity was increased in the mutant cells, including HSPCs. Our data thus indicate that the IRE1–XBP1 pathway is a potential therapeutic target for the eradication of the mutant clone in uncommitted HSPCs in patients with *CALR*-mutated myeloproliferative neoplasms.

In conclusion, high-throughput linking of single-cell genotypic and transcriptomic data underscored the dependency of somatic mutations on cell identity in human haematopoiesis, and enabled us to superimpose the native differentiation tree with a tree corrupted by somatic mutations. GoT further provides a means to gain insight into the integration of clonal diversification with lineage plasticity[47] or differentiation topologies[48] across cancer. Thus, GoT may pave the way to resolving central questions related to the link between genetic mutations and cellular identities, and help to unravel the underlying programs that enable clonal expansions and evolution in human neoplasms.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1367-0.

Received: 24 October 2018; Accepted: 5 June 2019;
Published online 3 July 2019.

1. Sperling, A. S., Gibson, C. J. & Ebert, B. L. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat. Rev. Cancer* **17**, 5–19 (2017).
2. Landau, D. A. et al. The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. *Nat. Commun.* **8**, 2185 (2017).
3. Burger, J. A. et al. Clonal evolution in patients with chronic lymphocytic leukaemia developing resistance to BTK inhibition. *Nat. Commun.* **7**, 11589 (2016).
4. Landau, D. A. et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* **152**, 714–726 (2013).
5. Landau, D. A. et al. Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–530 (2015).
6. Nangalia, J. et al. Somatic *CALR* mutations in myeloproliferative neoplasms with nonmutated *JAK2*. *N. Engl. J. Med.* **369**, 2391–2405 (2013).
7. Klampfl, T. et al. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N. Engl. J. Med.* **369**, 2379–2390 (2013).
8. Giustacchini, A. et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. *Nat. Med.* **23**, 692–702 (2017).
9. Cheow, L. F. et al. Single-cell multimodal profiling reveals cellular epigenetic heterogeneity. *Nat. Methods* **13**, 833–836 (2016).
10. Macosko, E. Z. et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
11. Klein, A. M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201 (2015).
12. Elf, S. et al. Mutant calreticulin requires both its mutant C-terminus and the thrombopoietin receptor for oncogenic transformation. *Cancer Discov.* **6**, 368–381 (2016).
13. Defour, J. P., Chachoua, I., Pecquet, C. & Constantinescu, S. N. Oncogenic activation of MPL/thrombopoietin receptor by 17 mutations at W515: implications for myeloproliferative neoplasms. *Leukemia* **30**, 1214–1216 (2016).
14. Kollmann, K. et al. A novel signalling screen demonstrates that CALR mutations activate essential MAPK signalling and facilitate megakaryocyte differentiation. *Leukemia* **31**, 934–944 (2017).
15. Marty, C. et al. Calreticulin mutants in mice induce an MPL-dependent thrombocytosis with frequent progression to myelofibrosis. *Blood* **127**, 1317–1324 (2016).
16. Nivarthi, H. et al. Thrombopoietin receptor is required for the oncogenic function of CALR mutants. *Leukemia* **30**, 1759–1763 (2016).
17. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).
18. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
19. Lopez, R., Regier, J., Cole, M. B., Jordan, M. I. & Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nat. Methods* **15**, 1053–1058 (2018).
20. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
21. Farrell, J. A. et al. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar313 (2018).
22. Chen, E. et al. Distinct clinical phenotypes associated with *JAK2V617F* reflect differential STAT1 signaling. *Cancer Cell* **18**, 524–535 (2010).
23. Rampal, R. et al. Integrated genomic analysis illustrates the central role of JAK–STAT pathway activation in myeloproliferative neoplasm pathogenesis. *Blood* **123**, e123–e133 (2014).
24. Tirosh, I. et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).
25. Georgantas, R. W. III et al. Microarray and serial analysis of gene expression analyses identify known and novel transcripts overexpressed in hematopoietic stem cells. *Cancer Res.* **64**, 4434–4441 (2004).
26. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
27. Hetz, C. & Papa, F. R. The unfolded protein response and cell fate control. *Mol. Cell* **69**, 169–181 (2018).
28. Pronier, E. et al. Targeting the CALR interactome in myeloproliferative neoplasms. *JCI Insight* **3**, e122703 (2018).
29. Lau, W. W., Hannah, R., Green, A. R. & Göttgens, B. The JAK–STAT signaling pathway is differentially activated in *CALR*-positive compared with *JAK2V617F*-positive ET patients. *Blood* **125**, 1679–1681 (2015).
30. Shivarov, V., Ivanova, M. & Tiu, R. V. Mutated calreticulin retains structurally disordered C terminus that cannot bind Ca^{2+}: some mechanistic and therapeutic implications. *Blood Cancer J.* **4**, e185 (2014).
31. Zini, R. et al. *CALR* mutational status identifies different disease subtypes of essential thrombocythemia showing distinct expression profiles. *Blood Cancer J.* **7**, 638 (2017).
32. Wu, J. et al. ATF6α optimizes long-term endoplasmic reticulum function to protect cells from chronic stress. *Dev. Cell* **13**, 351–364 (2007).
33. van Galen, P. et al. The unfolded protein response governs integrity of the haematopoietic stem-cell pool during stress. *Nature* **510**, 268–272 (2014).
34. Lee, A. H., Iwakoshi, N. N. & Glimcher, L. H. XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. *Mol. Cell. Biol.* **23**, 7448–7459 (2003).
35. Cubillos-Ruiz, J. R. et al. ER stress sensor XBP1 controls anti-tumor immunity by disrupting dendritic cell homeostasis. *Cell* **161**, 1527–1538 (2015).
36. Yoshida, H., Matsui, T., Yamamoto, A., Okada, T. & Mori, K. XBP1 mRNA is induced by ATF6 and spliced by IRE1 in response to ER stress to produce a highly active transcription factor. *Cell* **107**, 881–891 (2001).
37. Luo, J. L., Kamata, H. & Karin, M. IKK/NF-κB signaling: balancing life and death—a new approach to cancer therapy. *J. Clin. Invest.* **115**, 2625–2632 (2005).
38. Stein, S. J. & Baldwin, A. S. Deletion of the NF-κB subunit p65/RelA in the hematopoietic compartment leads to defects in hematopoietic stem cell function. *Blood* **121**, 5015–5024 (2013).
39. Abu-Zeinah, G. et al. Myeloproliferative neoplasm (MPN) driver mutations are enriched during hematopoietic stem cell differentiation in patterns that correlate with clinical phenotype and treatment response. *Blood* **132**, 4317 (2018).
40. Castro-Malaspina, H., Rabellino, E. M., Yen, A., Nachman, R. L. & Moore, M. A. Human megakaryocyte stimulation of proliferation of bone marrow fibroblasts. *Blood* **57**, 781–787 (1981).
41. Ciurea, S. O. et al. Pivotal contributions of megakaryocytes to the biology of idiopathic myelofibrosis. *Blood* **110**, 986–993 (2007).
42. Terui, T. et al. The production of transforming growth factor-beta in acute megakaryoblastic leukemia and its possible implications in myelofibrosis. *Blood* **75**, 1540–1548 (1990).
43. Obeng, E. A. et al. Physiologic expression of *Sf3b1*^{K700E} causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. *Cancer Cell* **30**, 404–417 (2016).
44. Saikia, M. et al. Simultaneous multiplexed amplicon sequencing and transcriptome profiling in single cells. *Nat. Methods* **16**, 59–62 (2019).
45. Hill, A. J. et al. On the design of CRISPR-based single-cell molecular screens. *Nat. Methods* **15**, 271–274 (2018).
46. Kleppe, M. et al. Dual targeting of oncogenic activation and inflammatory signaling increases therapeutic efficacy in myeloproliferative neoplasms. *Cancer Cell* **33**, 785–787 (2018).
47. Mu, P. et al. *SOX2* promotes lineage plasticity and antiandrogen resistance in *TP53*- and *RB1*-deficient prostate cancer. *Science* **355**, 84–88 (2017).
48. Suvà, M. L. et al. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580–594 (2014).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment.

Species-mixing experiment. Previously published UT7 and Ba/F3 cell lines that express human *MPL* and either human wild-type *CALR* or mutant *CALR* (type 1, 52-bp deletion), provided by the laboratory of A. Mullally, were used for the species-mixing study[12]. In brief, human *MPL*-expressing Ba/F3 and UT7 cell lines were generated by retroviral transduction, after which they were subjected to infection with *CALR* variant lentiviral supernatants. Wild-type UT7 cells and mutant Ba/F3 cells were mixed in equal proportions and underwent GoT, targeting ~1,000 cells. Although UT7 has been listed as a commonly misidentified cell line, it was used for the sole purpose of validating the *CALR* mutation status of the cells. All cell lines used in the study were tested for mycoplasma contamination.

Patient samples. The study was approved by the local ethics committee and by the Institutional Review Board of Memorial Sloan Kettering Cancer Center and Weill Cornell Medicine, and conducted in accordance with the Declaration of Helsinki protocol. All patients provided informed consent. Cryo-preserved bone marrow mononuclear cells or peripheral blood mononuclear cells from patients with documented *CALR* mutations were retrieved after a database search (see Supplementary Table 1 for clinical information). Cryopreserved bone marrow mononuclear cells or peripheral blood mononuclear cells were thawed and stained using standard procedures (10 min, 4 °C) with the surface antibody CD34-PE-Vio770 (clone AC136, lot no. 5180718070, dilution 1:50, Miltenyi Biotec) and DAPI (Sigma-Aldrich). Cells were then sorted for DAPI$^-$, CD34$^+$ and DAPI$^-$, CD34$^-$ cells using BD Influx at the Weill Cornell Medicine flow cytometry core.

Targeted myeloid panel. To identify recurrent somatic mutations and their VAF in patient samples, targeted next-generation sequencing was performed on DNA samples extracted from unfractionated peripheral blood mononuclear cells (patients ET09, MF01, MF02, MF03 and MF04), CD34$^-$ sorted bone marrow mononuclear cells (patients ET02, ET03, ET04 and ET05), CD34$^+$ sorted bone marrow mononuclear cells (patient ET01) and CD34$^+$ sorted peripheral blood mononuclear cells (patient MF05), as previously described[49]. In brief, the targeted enrichment of 45 genes (*ABL1*, *ASXL1*, *BCOR*, *BRAF*, *CALR*, *CBL*, *CEBPA*, *DNMT3A*, *ETV6*, *EZH2*, *FAM5C*, *FLT3*, *GATA1*, *GATA2*, *HNRNPK*, *IDH1*, *IDH2*, *IKZF1*, *JAK1*, *JAK2*, *KDM6A*, *KIT*, *KRAS*, *MPL*, *NFE2*, *NOTCH1*, *NPM1*, *NRAS*, *PHF6*, *PTPN11*, *RAD21*, *RUNX1*, *SETBP1*, *SF3B1*, *SH2B3*, *SMC1A*, *SMC3*, *SRSF2*, *STAG2*, *SUZ12*, *TET2*, *TP53*, *U2AF1* and *ZRSR2*) that are recurrently mutated in myeloid malignancies was performed using the Thunderstorm system (Raindance Technologies) with a custom primer panel followed by sequencing using the Illumina MiSeq (v.3 chemistry).

GoT. Extending recent experience with targeted amplicon sequencing in scRNA-seq[44,45], we developed GoT to simultaneously capture genotyping data and whole transcriptomic data in single cells by adapting the 10x Genomics platform. The standard 10x Genomics Chromium 3′ (v.2 or v.3 chemistry) and 5′ libraries were carried out according to the manufacturer's recommendations until after emulsion breakage and recovery of first strand cDNA (Fig. 1a, step 1). For 3′ libraries, if the targeted gene of interest (for example, *SF3B1*) was not robustly detected by the standard 10x procedure (that is, if <60% of the expected cells showed expression), on the basis of a priori knowledge in a similar dataset a gene-specific primer was spiked into 10x primer mix at 1% of the concentration of the cDNA amplification primers for the initial cDNA PCR step (Fig. 1a; see Supplementary Table 5 for list of primers and Extended Data Fig. 1b, c for primer positions). For 5′ libraries, the presence of 10x cell barcodes and UMIs on the 3′ side of the transcript enabled a gene-specific primer spike-in during the reverse transcription (RT) step (guide RT primer, 0.12 μM final concentration, Supplementary Table 5) to increase capture and detection of the transcript of interest (for example, *JAK2*). At the cDNA amplification step, another spike-in primer (additive primer) is added to increase the yield of the same transcript. During the amplification step, for 3′ libraries v.2 chemistry only, the 10x cDNA library underwent an extra cycle of PCR beyond the manufacturer's recommended number of cycles. (3′ v.3 chemistry and 5′ libraries do not require extra cycles of PCR at the amplification step.) After cDNA amplification and clean-up with SPRIselect, a small portion of the cDNA library (3 μl for 3′ v.2 and 10 μl for 3′ v.3 chemistry and 5′ libraries) was aliquoted for targeted genotyping, and the remaining cDNA underwent the standard 10x protocol. In the case of 3′ v.2 chemistry, the cDNA set aside for GoT was amplified for 3 to 4 additional cycles using KAPA HiFi HotStart ReadyMix (KAPABiosystems) and 10x primer mix to provide sufficient material for the enrichment step. After clean-up, locus-specific reverse primers and generic forward sample index (SI)-PCR oligonucleotide (10x Genomics) were used to amplify the site of interest of the cDNA template (Extended Data Fig. 1b, c, Supplementary Table 5). The number of PCR cycles was determined experimentally and was dependent on the level of expression of the targeted gene (for example, ten cycles were used for *CALR*). The locus-specific reverse primers contain a partial Illumina read 2 handle, a stagger to increase the complexity of the library for optimal sequencing and a gene-specific region to enable specific priming. The SI-PCR oligonucleotide anneals to the partial Illumina read 1 sequence at the 3′ or 5′ end of the molecule when using 3′ or 5′ libraries, respectively, which preserves the cell barcode and UMI (Extended Data Fig. 1b, c). After the initial amplification and solid phase reversible immobilization (SPRI) purification to remove unincorporated primers, a second PCR was performed with a generic forward PCR primer (P5_generic) to retain the cell barcode and UMI, together with an RPI-x primer (Illumina) to complete the P7 end of the library and add a sample index. The targeted amplicon library was subsequently spiked into the remainder of the 10x library to be sequenced together on a HiSeq 2500 or sequenced separately on MiSeq (Illumina). The cycle settings were as follows: 26 cycles for read 1, 98 or 130 cycles for read 2, and 8 cycles for sample index for 3′ v.2 chemistry and 5′ libraries; or 28 cycles for read 1, 98 or 130 cycles for read 2, and 8 cycles for sample index for 3′ v.3 chemistry.

Circularization GoT. For patient samples, we used the same starting material as for GoT (that is, non-fragmented 10x cDNA fraction); for the *JAK2* cDNA mixing study, we mixed barcoded cDNA from two cell lines (TF-1, *JAK2* wild type (ATCC CRL-2003); HEL, homozygous *JAK2*^{V617F} (ATCC TIB-180). With these cDNA libraries, we first performed a PCR to enrich for the amplicon, amplifying from about 50 bp upstream of our region of interest to the 3′ end of the 10x library fragment (therefore retaining the cell barcode and UMI), using KAPA HiFi Uracil+ master mix (Kapa Biosystems) and the following PCR conditions: 98 °C for 3 min; 10 to 20 cycles of 98 °C for 20 s, 65 °C for 30 s, 72 °C for 2 min and 72 °C for 5 min. Complementary U-overhang are added to the forward (Fw) and reverse (Rv) primers to allow circularization: Fw-primer no. 1: AGGUCAGTCU-[specific to approximately 50 bp upstream of the locus], Rv-primer no. 1: AGACUGACCUCTACACGACGCTCTTCCGATCT (Extended Data Fig. 1b, c, Supplementary Table 5). For genes that are represented at low levels in the cDNA library (such as *SF3B1*), we specifically pre-enriched the gene of interest by doing a PCR that targeted about 100 bp upstream of our region of interest to the 3′ end of the 10x library fragment, using KAPA HiFi Ready mix (Kapa Biosystems) and the following PCR conditions: 95 °C for 3 min; 20 cycles of 98 °C for 20 s, 65 °C for 30 s, 72 °C for 2 min and 72 °C for 5 min. PCR product resulting from the first single or double PCR was then cleaned up and concentrated using 1.3× SPRI beads. Next, amplicon cohesive ends were created using 40 U/ml USERII enzyme (M5508-NEB) digestion for 1 h at 37 °C in 1× CutSmart buffer. Reaction was stopped by incubating for 10 min at 65 °C. Relying on complementary overhangs at both ends of the amplicon, circularization was performed in a large volume (>1 ml) to favour intra-molecule ligation. The following reaction was set up and incubated overnight at 16 °C: USERII-digested amplicon, 2,000 U/ml T4 ligase (NEB), 1× CutSmart Buffer (NEB) and 1 mM ATP (Roche). Next, T4 DNA ligase was inactivated by incubating for 15 min at 70 °C. Then, unwanted unligated products were removed by adding 6 U of lambda exonuclease (NEB, M0262S) in the ligation mix and incubating for 30 min at 37 °C. Exonuclease was inactivated for 20 min at 65 °C. Ligated product was cleaned up and concentrated using 1.3× SPRI beads. A second PCR was set up to retain the locus of interest and barcodes on the same molecule, while removing the unwanted 3′ downstream region of the targeted region. PCR reaction was set up and performed as previously described, using the following primers: Fw-primer no. 2: AGGUCAGTCU-[specific to 3′ end locus], Rv-primer no. 2: AGACUGACCU-[specific to 10 bp downstream of locus].

After PCR no. 2, SPRI clean-up, USERII digestion, overnight T4 ligation and lambda exonuclease digestion were performed as described above. After the second ligation, the ligated product was again cleaned up and concentrated using 1.3× SPRI beads. To increase ligation efficiency during the circularization step and to reduce protocol duration (from 3 days to 1 day), we further improved ligation by using the Gibson assembly molecular cloning approach. Instead of U-overhang handles, complementary Gibson handles are added to the forward and reverse primers to allow circularization after PCR no. 1 and PCR no. 2 (Extended Data Fig. 1b, c, Supplementary Table 5). PCR no. 1 and PCR no. 2 are performed as described above for the U-overhang version of this protocol, but using KAPA HiFi Ready mix. Ligation was then performed over 1 h at 50 °C in a large volume (>1 ml, 1× CutSmart Buffer) and using 10 μl of Gibson master mix (NEB, E2611). Finally, to linearize the product of ligation, we performed a third PCR: Fw-primer no. 3: CCTTGGCACCCGAGAATTCCA-[specific to 10 bp upstream of the locus] Rv-primer no. 3: SI-PCR (10x Genomics). We used KAPA HiFi master mix (Kapa Biosystems) and the following PCR conditions: 95 °C for 3 min; 10 cycles of 98 °C for 20 s, 65 °C for 30 s and 72 °C for 30 s; 72 °C for 5 min. After SPRI purification, a final PCR was performed with a generic forward PCR primer (P5_generic) and an RPI-x primer (Illumina) to complete the P7 end of the library and add a sample index (95 °C for 3 min; 5 cycles of 98 °C for 20 s, 67 °C for 30 s and 72 °C for 30 s; 72 °C for 5 min). This method generates amplicons that retain the contiguity of the original molecules but are short enough to cluster effectively to be sequenced with standard parameters. The targeted amplicon library was subsequently sequenced using PE150 on MiSeq (Illumina).

scRNA-seq data processing, alignment, cell-type classification and clustering. 10x data were processed using Cell Ranger 2.1.0 with default parameters. Reads

were aligned to the human reference sequence GRCh38 or hg19, or to mouse reference mm10 (species-mixing experiment). The genomic region of interest for genotyping was examined to determine how many UMIs with the targeted sequence were present in the conventional 10x data (Fig. 1c, Extended Data Fig. 3a, b).

The Seurat package (v.3.0) was used to perform unbiased clustering of the $CD34^+$ sorted cells from patient samples[18,50]. In brief, for individual datasets, cells with UMI < 200 or UMI > 3 s.d. from the mean UMI, and mitochondrial gene percentage $> 10\%$, were filtered. The data were log-normalized using a scale factor of 10,000. Before clustering, the essential thrombocythaemia and myelofibrosis datasets were integrated and underwent batch-correction within Seurat, which implements canonical correlation analysis and the principles of mutual nearest neighbour. Recommended settings were used for the integration (30 canonical correlation vectors for canonical correlation analysis in the FindIntegrationAnchors function and 30 principal components for the anchor weighting procedure in IntegrateData function). For the datasets, potential confounders (for example, numbers of UMI per cell and the proportion of mitochondrial genes) were regressed out of the data before principal component analysis was performed using variable genes. The JackStraw method was used to determine the statistically significant principal components to be used for graph-based clustering. *t*-SNE was used to visualize the clusters. Clusters were manually assigned on the basis of differentially expressed genes using the FindAllMarkers function using default settings (using all genes that are detected in a minimum of 25% of cells in either of the 2 comparison sets as input, and log-transformed fold change of 0.25 as the threshold). Wilcoxon rank-sum test was applied to rank genes, with the top 10 differentially expressed genes per cluster presented in Extended Data Fig. 4b. We identified 19 distinct clusters in the integrated data for samples ET01–ET05, which were annotated according to previously identified marker genes[26] (*t*-SNE in Fig. 1d, and clustering heat map and *t*-SNE with representative marker genes in Extended Data Fig. 4b, c). Pseudotime analysis was performed using the Monocle R package (v.3.8) for individual datasets[20] and the URD package (v.1.0.2) for the integrated datasets[21]. Linear mixed effects analysis was performed using the lme4 package (v.1.2-1). For mutant-frequency analysis between HSPCs and MkPs (Fig. 1g), genotype status was defined as the fixed effect, and as random effects we used intercepts for individual patients (subjects) and iterative downsampling. For integrated analysis of pseudotime comparisons (Fig. 1i) and gene-module expression (for example, Fig. 3c), genotype status was entered as the fixed effect and subjects as random effects. *P* values were obtained by likelihood ratio tests of the full model with the fixed effect against the model without the fixed effect[51].

IronThrone GoT for processing targeted amplicon sequences and mutation calling. To ensure correct priming, targeted amplicon reads (read 2) were screened for the presence of the primer sequence and the expected intervening sequence between the primer and the start of the mutation site ('shared sequence'; for circularization GoT, PCR no. 2 forward and PCR no. 2 reverse primer sequences) (Extended Data Fig. 2a, b). Ninety per cent of the reads from the mixing study showed the expected primer and shared sequences. Subsequently, for reads that passed the priming step, the corresponding read 1 was screened for the presence of the 16-bp- or 18-bp-long cell barcode that matched the cell barcode in the whitelist provided by 10x Genomics (Extended Data Fig. 10). For cell barcode reads that were one Hamming distance away from the whitelisted cell barcodes, the probability that the observed barcode originated from the whitelisted cell barcode was calculated, taking into account the base quality score at the differing base. The whitelisted cell barcode with the highest probability was used to replace the observed cell barcode only if the probability exceeded 0.99. For the duplicate reads with the same cell barcode and UMI, the genotype (wild type versus mutant) of the UMI was assigned on the basis of majority rule in supporting reads, or according to the read with the highest base quality score (in the rare cases in which only two supporting, but discordant, reads were available).

The species-mixing study for *CALR* (type 1) mutation was used to further optimize the analytical assignment of genotypes to cells, to overcome technical sources of noise such as PCR errors, ambient mRNA and PCR recombination, which may accompany targeted amplification in scRNA-seq. A mean ($\pm$ s.d.) of 83.6 ($\pm$ 95.3) *CALR* UMIs were detected per cell in the amplicon data, with 52 ($\pm$ 16.3) reads per UMI. We integrated targeted amplicon measures including base quality, number of base pair mismatches and number of duplicate reads per UMI, and determined optimized parameters that maximize the number of genotyped cells and minimize genotype misassignment (Extended Data Fig. 2c–g). Setting thresholds for the minimum number of duplicate reads and maximum frequency of mismatches contributed substantially to filtering out misassigned reads that were probably due to technical errors (such as PCR recombinations). A combination of a threshold of 2 or more duplicate reads for a given UMI and a threshold of allowing less than or equal to 0.2 mismatch ratio improved correct assignment of cells, and maximized the number of included cells for analysis, and was adopted in the analysis here (Extended Data Fig. 2c). Results of the precision and recall analyses also affirmed this combination of thresholds for minimum duplicate reads and maximum mismatch ratio (Extended Data Fig. 2d). Moreover, given the high number of *CALR* transcripts in the cell lines and thus higher potential effect of PCR recombination, cells were assigned as wild type or mutant if >90% of *CALR* amplicon UMIs were wild type or mutant, respectively.

To further assess the effect of various parameters of the amplicon reads on the precision of mutation calling, we tested these parameters in a random-forest classification using the mixing study, as implemented in the R randomForest package (v. 4.6-14)[52] (Extended Data Fig. 2e–g). Mean decrease accuracy was determined as a measure of importance of each variable used for the calculation of splits in trees (Extended Data Fig. 2e). For each combination of mismatch ratio and duplicate thresholds, the random forest was run 100 times to find the optimal number of random variables used in each tree and the minimum out-of-bag error was selected (Extended Data Fig. 2f, g). This random-forest analysis also showed a minimum duplicate read threshold of 2 and maximum mismatch ratio threshold of 0.2 to be optimal for minimizing misassignments, and a relatively low contribution of additional quality metrics.

Moreover, the genotyping information is derived from transcribed molecules and may be affected by the capture of transcripts from wild-type and mutant alleles of heterozygous mutations in primary patient samples (in which the median targeted amplicon UMI count per cell was 5 ($\pm$4.45, median absolute deviation)). This may be due to incomplete sampling of the transcript pool or to transcriptional bursts[53], which leads to skewed transcript pools. Consequently, as the number of UMIs per cell increases, the likelihood of capturing a mutant transcript increases, which results in an apparently higher frequency of mutant cells. Thus, the number of mutant reads may be underestimated in cells with lower amplicon UMI counts. Nonetheless, the frequency of mutant cells (for example, 26% in sample ET01) as determined by GoT using all cells that contain at least one UMI yielded values that were similar to that determined by bulk DNA exon sequencing of *CALR* from $CD34^+$ cells (mutant-cell fraction of 30%, based on VAF of 0.15 in a diploid heterozygous mutation). Although the bulk of the downstream analyses between *CALR* mutant and wild-type cells used a threshold of two or greater genotyping amplicon UMIs, we systematically applied three approaches to exclude the effect of this confounder (that is, the expression level of the target gene) on the conclusions. First, to exclude the possibility that higher *CALR* expression in committed progenitors results in a greater ability to detect mutant alleles, and thereby in a higher mutant-cell frequency, we downsampled all cells to a single amplicon UMI before mutation calling and found that the increase in mutation frequencies in MkP compared with HSPC remained unchanged (Fig. 1f, g). Second, we explored the sensitivity of the difference between mutant and wild-type cells (for example, pseudotime or mutant-cell frequency) by increasing the minimal amplicon UMI threshold allowed for mutation calling, and demonstrated that this did not effect the central findings of this study (Extended Data Fig. 5a, b). Third, we explicitly modelled the effect of *CALR* amplicon UMI in multivariable models (generalized linear model using R statistics package v.3.5.1 (for example, pseudotime analysis)), in which the number of amplicon UMIs was included in the model alongside the mutation status (Extended Data Fig. 5c). We further note that the GoT procedure did not result in loss of genes or UMIs per cell in comparison to published data of $CD34^+$ selected cells from the standard 10x library[54] (Extended Data Fig. 3d). For *XBP1* splicing analyses, we required a cell to have at least one unspliced *XBP1* for inclusion in the analyses.

Deep generative model for single-cell analysis. We applied a previously published[19] deep generative modelling approach for the single-cell analysis of samples ET01–ET05[19] (Extended Data Fig. 4d, e). Using the scVI package, we trained a variational auto-encoder that takes as input a feature vector for each cell, consisting of transcript counts for the 256 genes with the highest s.d. across all samples as well as an indicator for batch identity. Using 90% of cells for training and holding back 10% for validation, these features are provided to the variational auto-encoder with 64-unit hidden layers in both the encoder and decoder modules, and a four-dimensional internal latent vector of Gaussian-distributed values that provide a more concise representation of biological variability between cells. *t*-SNE is applied to these vectors for visualization.

Analysis of mutant-cell frequency. For integrated analysis of samples ET01–ET05 or MF01–MF04, an equal number of cells from each sample ($n = 900$ for essential thrombocythaemia and $n = 400$ for myelofibrosis) were subsampled randomly. Genotyping amplicon UMIs were downsampled ($\times$100 iterations) to one per cell and mutant-cell frequency was determined for each cluster for either the integrated dataset or individual samples. This frequency was then divided by the total mutant-cell frequency across all progenitor subsets for each of the iterations (Figs. 1f, g, 4c, 5h, Extended Data Fig. 5d).

Differential expression and gene-set enrichment analysis. For gene-module analysis, the aggregate gene-expression levels of modules of genes that are involved in biological processes of interest (see complete list of genes for each module in Supplementary Table 2) were calculated as $\log_2$ of the ratio of UMIs in the gene module per 10,000 UMIs per cell. The gene modules have been previously published[22,24–26,32,33] (Supplementary Table 2). Differential gene-expression analysis between mutant and wild-type cells for each of the progenitor clusters for each patient was performed via the FindMarkers function within the Seurat package

using the logistic regression for differential gene expression[55], with variable genes as input and requiring expression in at least 10% of cells in either group. UMI was included as a latent variable. The differentially expressed genes were examined individually for each patient; they were also examined in combination for each cluster across the patients by combining the *P* values for the differentially expressed genes using Fisher's method, and performing a weighted average of the $\log_2$-transformed fold change (Supplementary Table 3). Genes that were differentially expressed with false discovery rate < 0.1 and $\log_2$-transformed fold change ≥ 0.2 were included for gene-set enrichment analysis. Hypergeometric test for gene-set enrichment analysis was performed using the gProfileR package (v.0.6.7)[56]. Multiple hypothesis testing correction was performed using the g:SCS algorithm, developed by the authors of the gProfileR package. Kyoto Encyclopedia of Genes and Genomes (KEGG), Reactome, GO:MF and GO:BP data sources were included in the analyses (Supplementary Table 4).

Comparison of mutant allelic fraction in whole-exome sequencing and RNA-seq. We compared the mutant allelic fractions between genomic DNA and RNA—estimated from whole-exome sequencing (WES) and RNA-seq data, respectively—in five cancer cohorts (breast invasive carcinoma, head and neck squamous cell carcinoma, kidney renal clear cell carcinoma, lung adenocarcinoma and stomach adenocarcinoma). For this analysis, we thank T.-M. Kim (Cancer Research Institute, College of Medicine, The Catholic University of Korea) for sharing the datasets curated for a previous study[57]. In brief, the datasets of each of the cancer cohorts were initially prepared with somatic mutation sets from The Cancer Genome Atlas portal (https://portal.gdc.cancer.gov/). Then, reference and alternative alleles for these mutations were counted in .bam files of WES and RNA-seq using SAMtools mpileup[58], and filtered for >10 coverage of reads. We then converted genomic coordinates of the datasets from hg19 to hg38 assembly. We identified the frequencies of somatic mutations in cancer samples from CosmicCodingMuts.vcf (v.86) in the Catalogue of Somatic Mutations in Cancer (COSMIC) database[59]. Then, we further annotated the variants as oncogene or tumour-suppressor genes[60], and as driver or passenger mutations[61], using previously published definitions.

Determination of distance of targeted loci from 3′ or 5′ ends of transcripts. To identify Ensembl transcript identifiers that correspond to each mutation in the datasets of the five cancer cohorts described above, we matched them with COSMIC identifiers and annotated from the file of CosmicMutantExport.tsv (v.86). We used the biomaRt R package[62] with the GRCh38 version to annotate the transcript, including the length of transcript and the position of cDNA start codon in the transcript. The positions of the 5′ untranslated region ends were determined to calculate the distance from 5′ end to target site.

Oxford Nanopore Technology. The cDNA amplicon samples were barcoded by ONT 1D native barcoding kit EXP-NBD104. The barcoded samples were fed into the ONT SQK-LSK109 library preparation and sequencing workflow. FLO-MIN106 RevD flowcells and GridION X5 sequencer were used for sequencing. Data were base-called by ONT Guppy 2.3.1. For analysis, the adaptor sequences were trimmed using Porechop (https://github.com/rrwick/Porechop). Then, the reads were assessed for correct priming as shown in Extended Data Fig. 9d. The correctly primed reads were aligned to the reference genome (Grch38) with minimap2[63] (v.2.16) for variant calling. The cell barcodes underwent the same processing as described above for IronThrone GoT (Extended Data Figs. 9d, 10).

ddPCR. Peripheral blood from three patients with essential thrombocythaemia with mutations in *CALR* underwent Ficoll density gradient separation, immunomagnetic selection for $CD34^+$ cells (Miltenyi Biotech) and FACS (Influx, Becton Dickinson) using PeCy7-labelled CD34, clone 561 (lot no. B257238, BioLegend), APC-labelled CD38, clone HIT2 (lot no. B247250, BioLegend) and FITC-labelled CD10, clone HI10a (lot no. B254556, BioLegend) antibodies were used to isolate $CD34^+CD38^-$, $CD34^+CD38^+$ and $CD34^+CD10^+$ cell compartments. DNA was extracted from sorted cells (Qiagen) and the VAF of *CALR* mutations was measured by ddPCR (QX200 Droplet Digital PCR System, Bio-Rad) with primers that specifically detect *CALR* type 1 mutations (52-bp deletion (p.L367fs*46), *CALR* type 2 mutations (5-bp TTGTC insertion (p.K385fs*47) or wild-type alleles.

Single-cell colony genotyping assay. Viably frozen mononuclear cells were thawed and plated in H4434 Methocultä medium (StemCell Technologies) containing recombinant human SCF, GM-CSF, IL-3 and EPO according to the manufacturer's specifications. Individual colonies ($n = 94$) were picked from the methylcellulose medium after 14 days of culture at 37 °C and sequenced by Sanger sequencing for *SF3B1*, *CALR* and *NFE2* mutations using primers listed in Supplementary Table 5.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All of the sequencing data are available via the Gene Expression Omnibus (GEO) under the accession number GSE117826. Any other relevant data are available from the corresponding author upon reasonable request.

Code availability

The IronThrone GoT pipeline is available on GitHub at https://github.com/landau-lab/IronThrone-GoT.

49. Geyer, J. T. et al. Oligomonocytic chronic myelomonocytic leukemia (chronic myelomonocytic leukemia without absolute monocytosis) displays a similar clinicopathologic and mutational profile to classical chronic myelomonocytic leukemia. *Mod. Pathol.* **30**, 1213–1222 (2017).
50. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
51. Bolker, B. M. et al. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**, 127–135 (2009).
52. Liaw, A. & Wiener, M. Classification and regression by randomForest. *R News* **2**, 18–22 (2002).
53. Reinius, B. & Sandberg, R. Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat. Rev. Genet.* **16**, 653–664 (2015).
54. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8**, 14049 (2017).
55. Ntranos, V., Yi, L., Melsted, P. & Pachter, L. A discriminative learning approach to differential expression analysis for single-cell RNA-seq. *Nat. Methods* **16**, 163–166 (2019).
56. Reimand, J. et al. g:Profiler–a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89 (2016).
57. Rhee, J. K., Lee, S., Park, W. Y., Kim, Y. H. & Kim, T. M. Allelic imbalance of somatic mutations in cancer genomes and transcriptomes. *Sci. Rep.* **7**, 1653 (2017).
58. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
59. Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* **45**, D777–D783 (2017).
60. Vogelstein, B. et al. Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
61. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **174**, 1034–1035 (2018).
62. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protocols* **4**, 1184–1191 (2009).
63. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).

Acknowledgements The work was enabled by the Weill Cornell Epigenomics Core and Flow Cytometry Core. We thank A. Mullally (Brigham and Women's Hospital) for sharing the cell lines for the species-mixing study, and N. Kuchine (Weill Cornell Medicine) for helping us to acquire one of the patient samples. R.C. is supported by Lymphoma Research Foundation and Marie Skłodowska-Curie fellowships. R.M.M. is supported by a Medical Scientist Training Program grant from the National Institute of General Medical Sciences of the National Institutes of Health, awarded to the Weill Cornell, Rockefeller University and Memorial Sloan Kettering Cancer Center Tri-Institutional MD-PhD Program (T32GM007739). G.A.-Z. and J.M.S. are supported by Cancer Research & Treatment Fund (CR&T). J.R.C.-R. is supported by the Stand Up to Cancer Innovative Research Grant (SU2C-AACR-IRG-03-16) and Department of Defense Early-Career Investigator Award (W81XWH-16-1-0438). D.A.L. is supported by the Burroughs Wellcome Fund Career Award for Medical Scientists, the American Society of Hematology Scholar Award, Pershing Square Sohn Prize for Young Investigators in Cancer Research and the National Institutes of Health Director's New Innovator Award (DP2-CA239065). This work was also supported by the Leukemia Lymphoma Society Translational Research Program, Columbia University Physical Sciences in Oncology Center Pilot Grant (U54CA193313), National Heart Lung and Blood Institute (R01HL145283-01) and Stand Up To Cancer Innovative Research Grant (SU2C-AACR-IRG-0616). Stand Up To Cancer is a program of the Entertainment Industry Foundation. Research grants are administered by the American Association for Cancer Research, the scientific partner of SU2C.

Author contributions A.S.N., K.-T.K., R.C., P.S., O.A.-W. and D.A.L. devised the research strategy. A.S.N., K.-T.K., R.C., P.S. and D.A.L. developed the tools. A.S.N., R.C., P.S., C.A., N.D.O., A.A., C.S., M.M., J.T., X.D., R.M.M., E.H. and G.A.-Z. performed the experiments. A.S.N., K.-T.K., R.B., A.P. and F.I. performed the analyses. A.S.N., K.-T.K., R.C., P.S. and D.A.L. wrote the manuscript. A.S.N., K.-T.K., R.C., J.T., P.S., J.R.C.-R., W.T., R.H., J.M.S., R.R., O.A.-W. and D.A.L. helped to interpret results. All authors reviewed and approved the final manuscript.

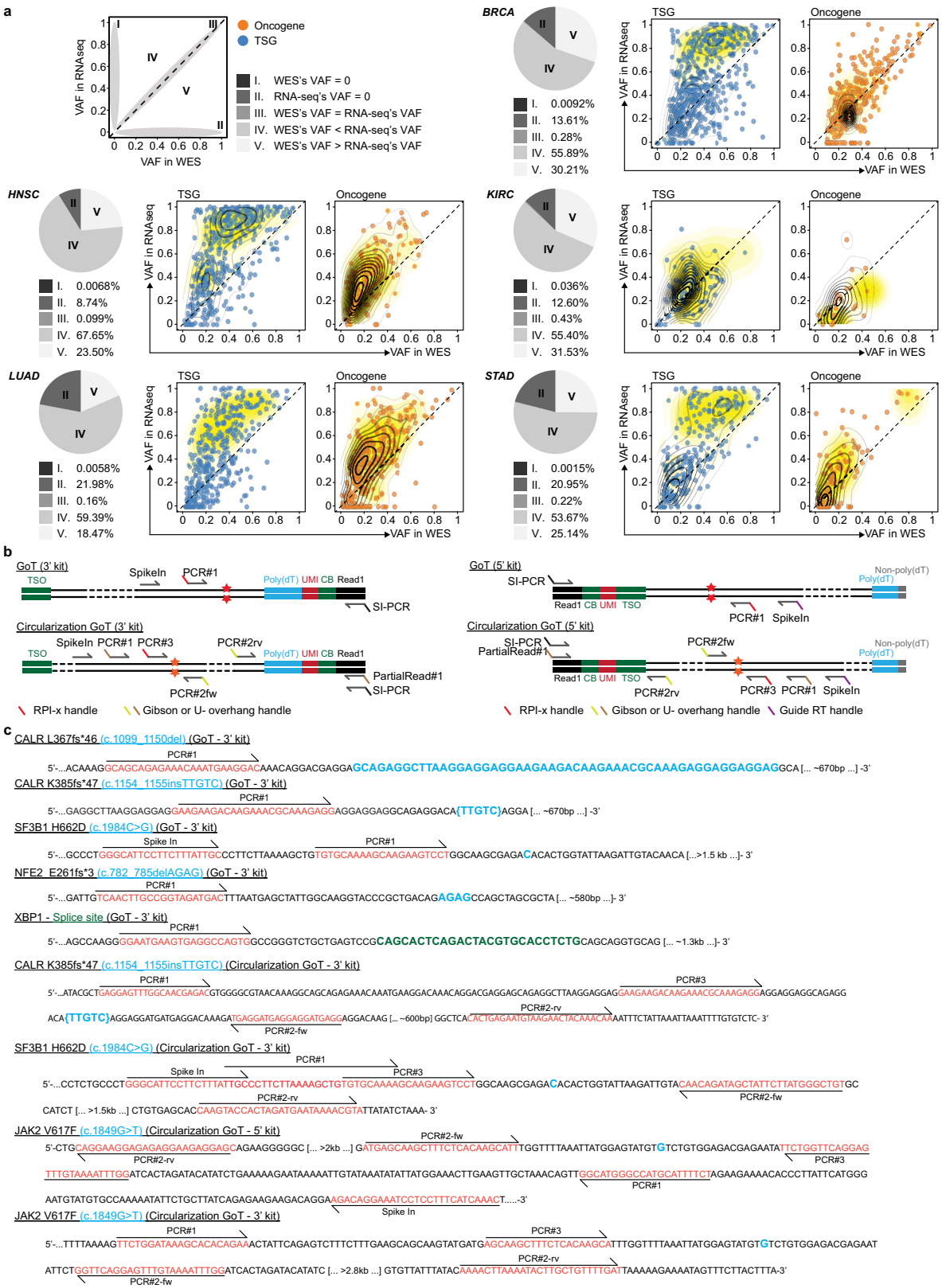
Competing interests The authors declare no competing interests.

Additional information

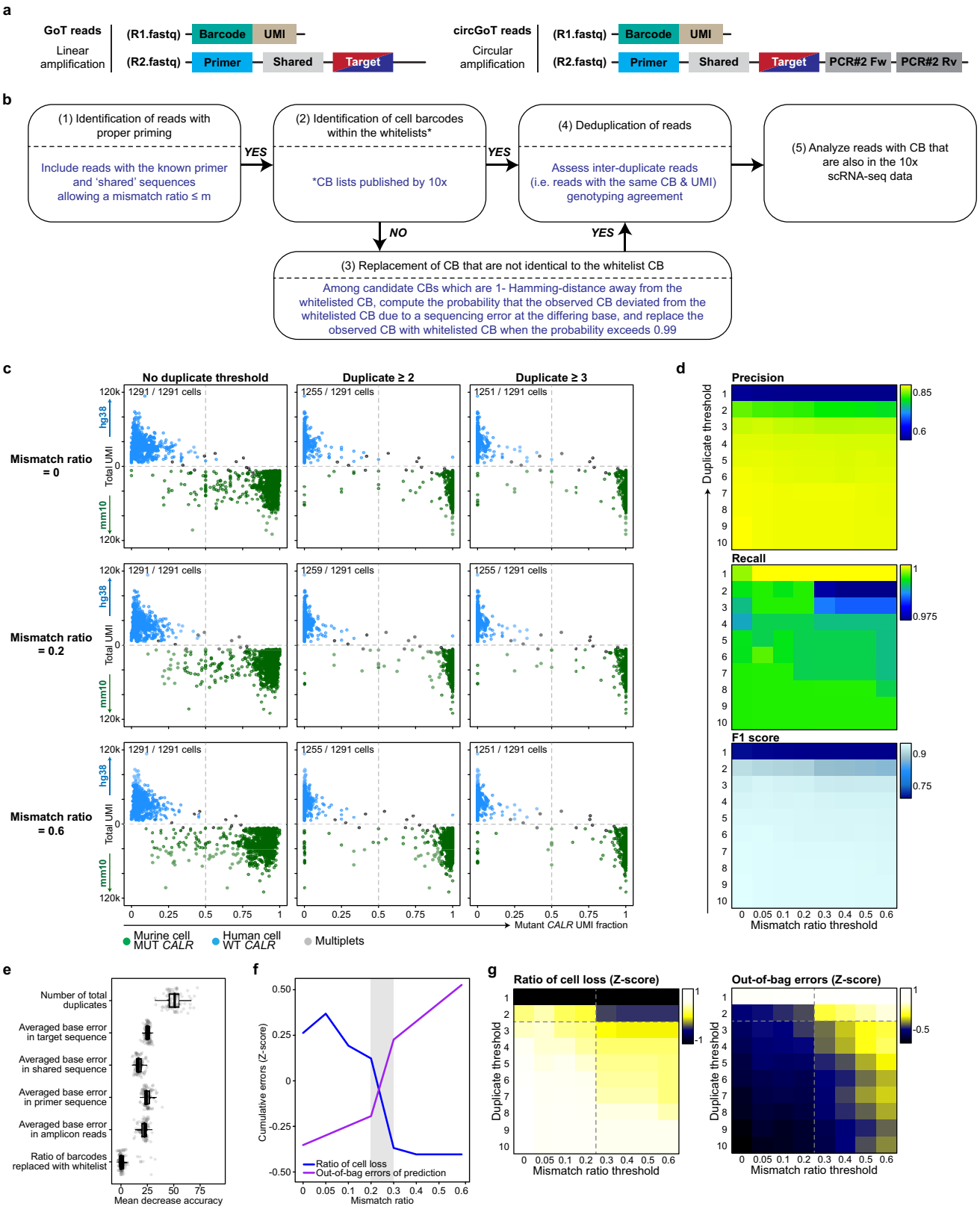
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1367-0.

Correspondence and requests for materials should be addressed to D.A.L.

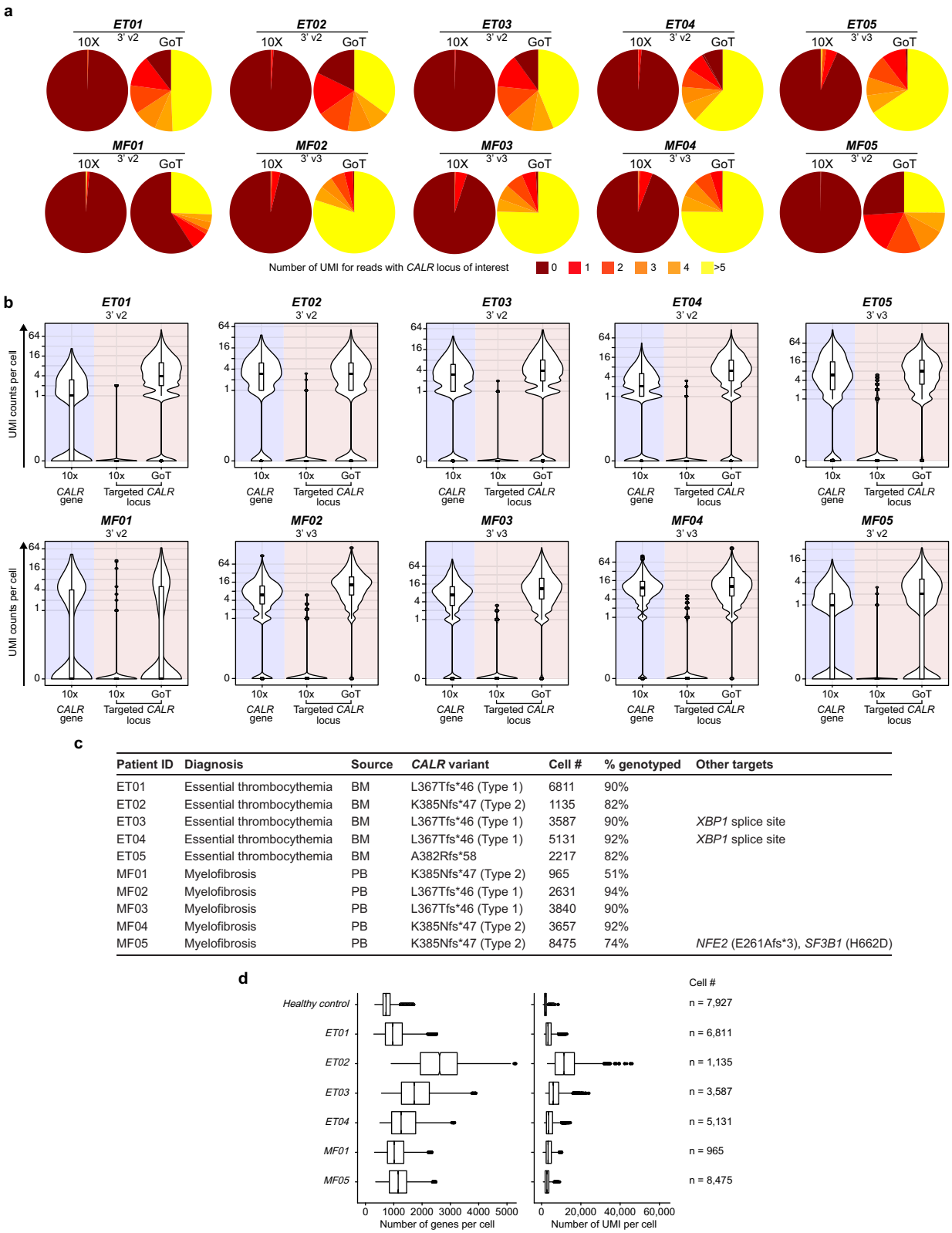
Peer review information *Nature* thanks Benjamin Lamarck Emert, Davis McCarthy, Arjun Raj and the other anonymous reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | Comparison of VAF between WES and RNA-seq, and primer sequences and positions of linear and circularization GoT. a, Pie charts show the fraction of variants, which are categorized as described in the top panel. The distribution of the mutant allele fraction is annotated as oncogene or tumour-suppressor gene (TSG) (according to previously published definitions[60,61]). Diagonal dashed lines indicate an equal allelic fraction between WES and RNA-seq. Yellow density contours represent driver distributions. BRCA, breast invasive carcinoma; HNSC, head and neck squamous cell carcinoma; KIRC, kidney renal clear cell carcinoma; LUAD, lung adenocarcinoma; STAD, stomach adenocarcinoma. **b**, Schematic localization of primers for linear GoT and circularization GoT for 3′ and 5′ libraries. **c**, Primer positions and sequences of the regions targeted by GoT and circularization GoT.

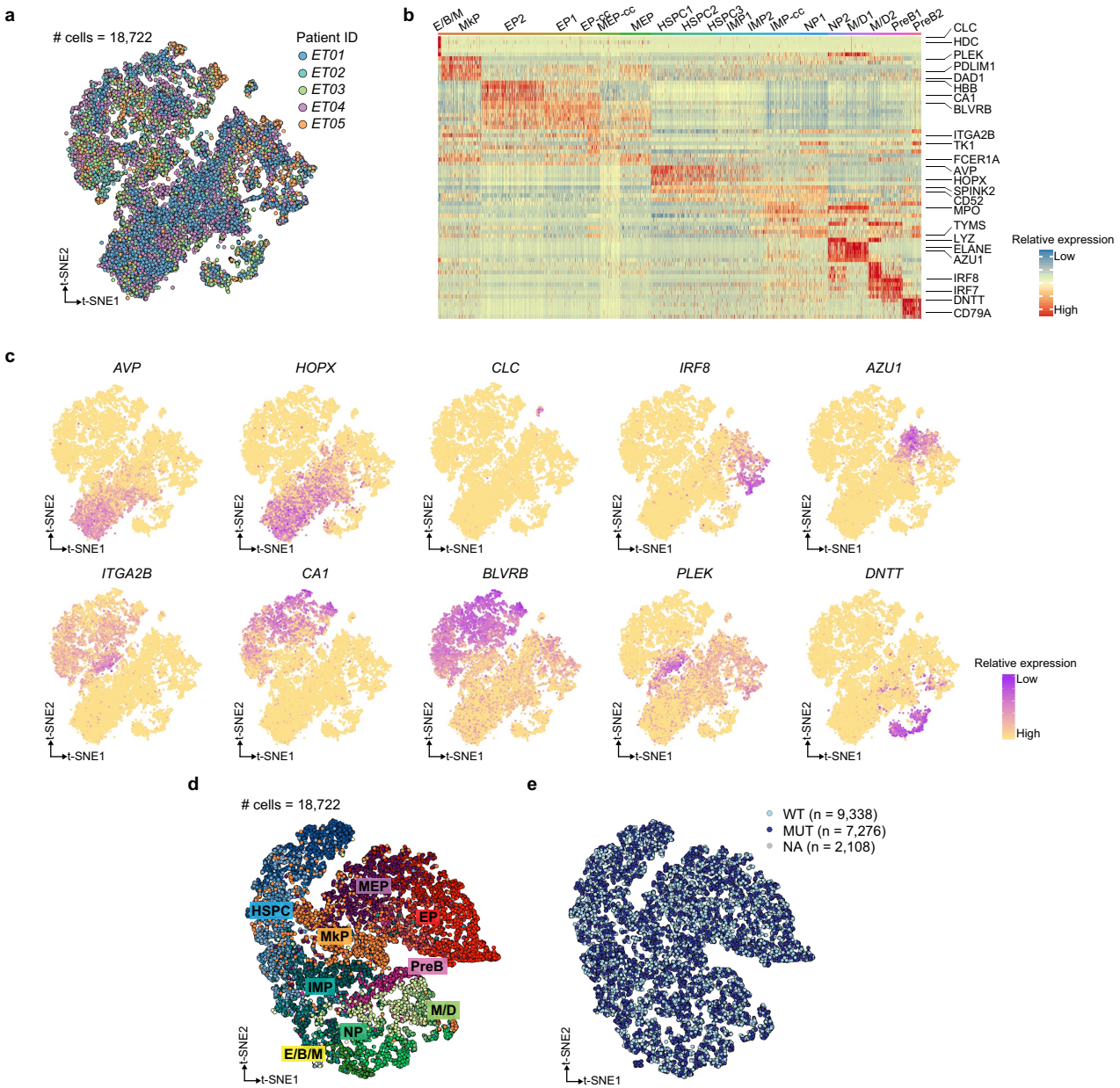


Extended Data Fig. 2 | Optimization of parameters in processing targeted amplicon sequences in the IronThrone GoT pipeline. **a**, Representation of amplicon reads. **b**, Flow chart of the GoT analysis pipeline (Methods). CB, cell barcode. **c**, Mouse (green) and human (blue) genome alignment of 10x data (*y* axes) with genotyping data by GoT (*x* axes) with various thresholds for minimum duplicate reads (across) and maximum mismatch ratio (down). **d**, Results of precision, recall and F_1 score analysis for combinations of minimum duplicate reads and maximum mismatch ratios. **e**, Measure of the importance of each variable used for the calculation of splits in trees in random-forest classification test. **f**, Ratio of cell loss and genotyping errors (*z*-score on *y* axis) based on mismatch ratio thresholds (*x* axis). The area of intersection is highlighted in grey around the mismatch ratio 0.2. **g**, Heat maps showing *z*-scores of the number of filtered cells (left) and predicted error rates (right) from random-forest classification tests for combinations of minimum duplicate reads and maximum mismatch ratio thresholds.

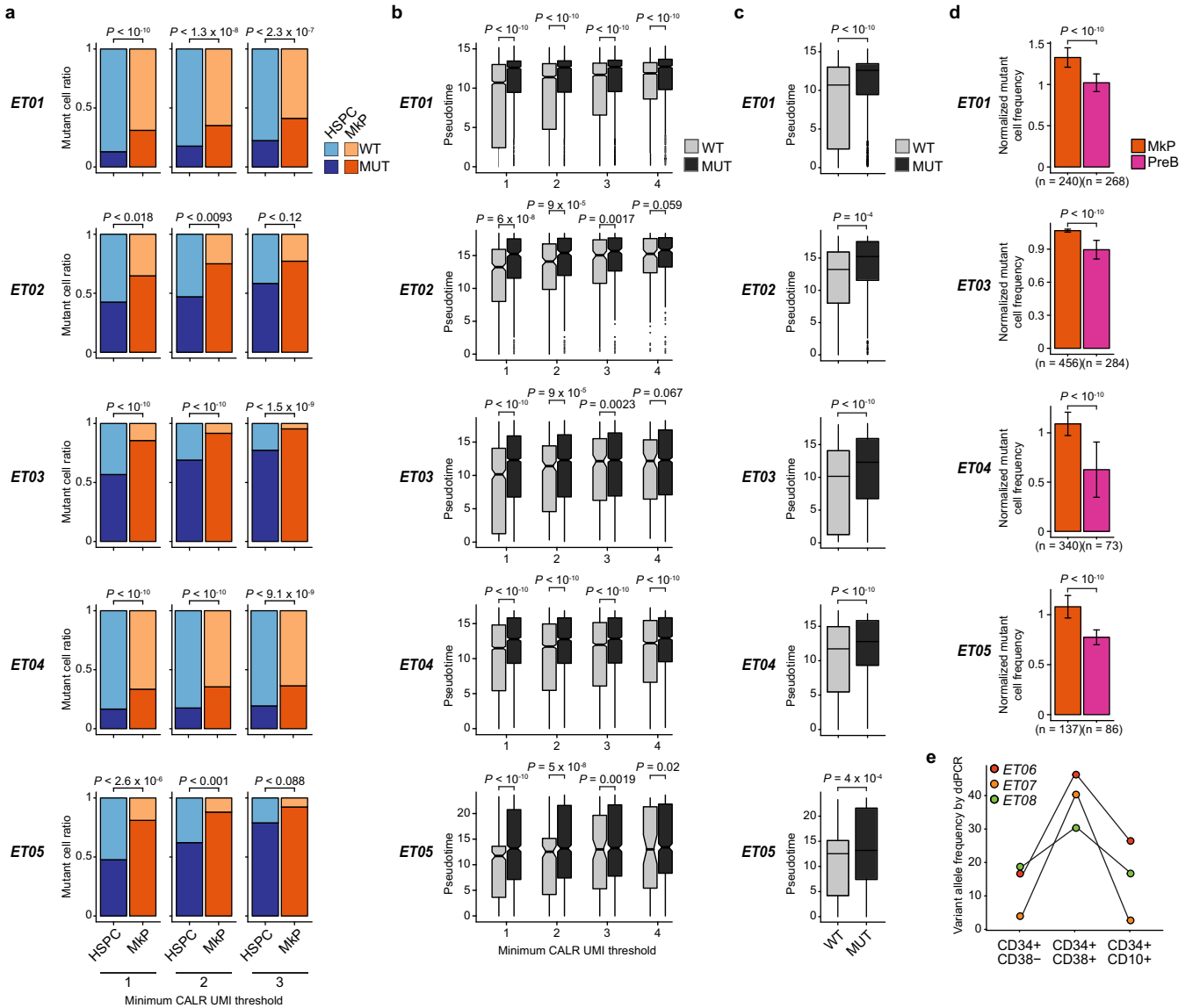


Patient ID	Diagnosis	Source	*CALR* variant	Cell #	% genotyped	Other targets
ET01	Essential thrombocythemia	BM	L367Tfs*46 (Type 1)	6811	90%	
ET02	Essential thrombocythemia	BM	K385Nfs*47 (Type 2)	1135	82%	
ET03	Essential thrombocythemia	BM	L367Tfs*46 (Type 1)	3587	90%	*XBP1* splice site
ET04	Essential thrombocythemia	BM	L367Tfs*46 (Type 1)	5131	92%	*XBP1* splice site
ET05	Essential thrombocythemia	BM	A382Rfs*58	2217	82%	
MF01	Myelofibrosis	PB	K385Nfs*47 (Type 2)	965	51%	
MF02	Myelofibrosis	PB	L367Tfs*46 (Type 1)	2631	94%	
MF03	Myelofibrosis	PB	L367Tfs*46 (Type 1)	3840	90%	
MF04	Myelofibrosis	PB	K385Nfs*47 (Type 2)	3657	92%	
MF05	Myelofibrosis	PB	K385Nfs*47 (Type 2)	8475	74%	*NFE2* (E261Afs*3), *SF3B1* (H662D)

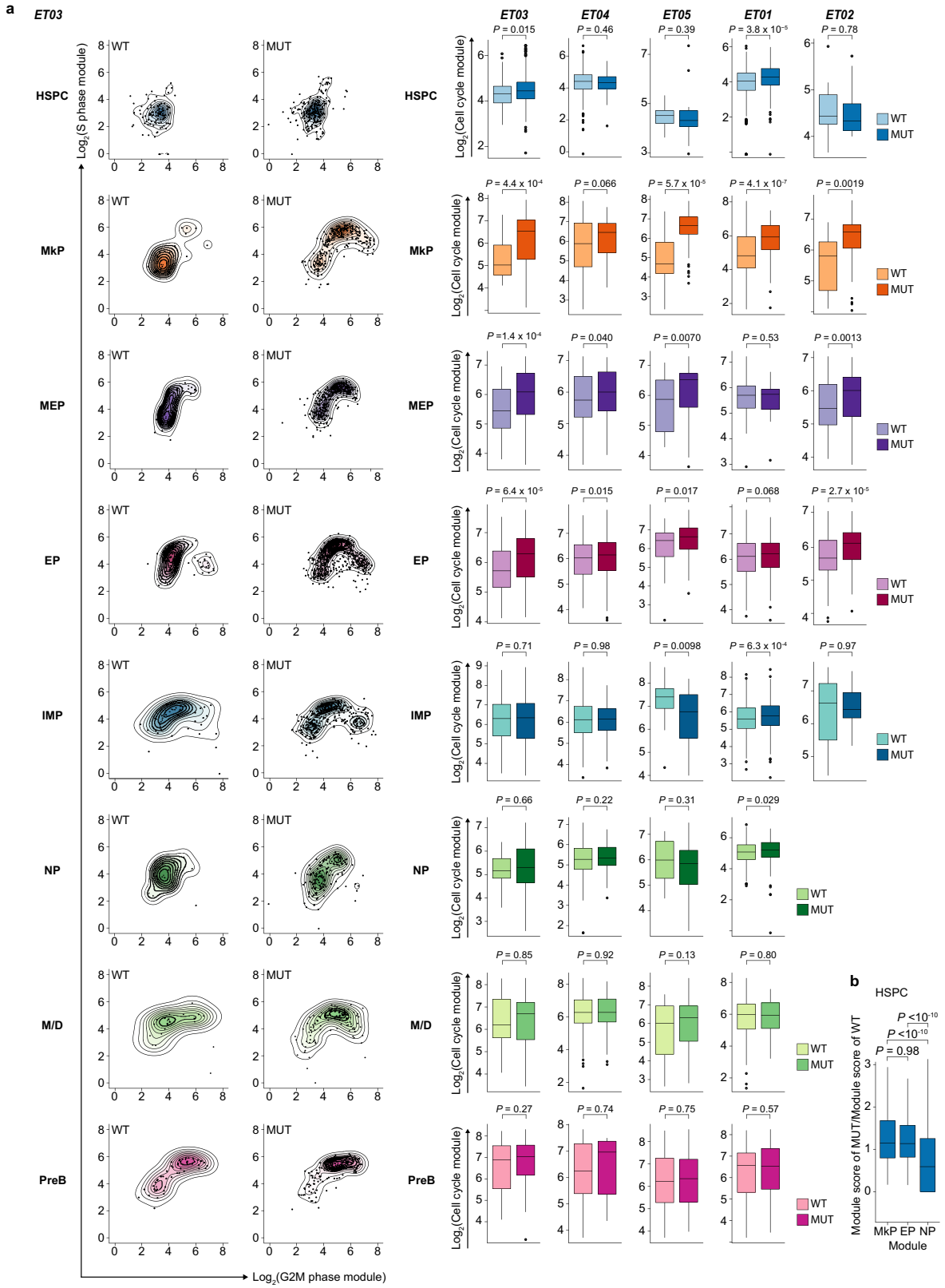
Extended Data Fig. 3 | GoT captures genotyping information of single cells through cDNA. a, Percentage of cells by number of UMIs with the *CALR*-mutation locus captured in standard 10x data (left panels) and GoT data (right panels) (see **c** for cell numbers in each sample). **b**, Number of UMIs per cell of *CALR* transcript from standard 10x data (blue shading) or targeted *CALR* locus from standard 10x or GoT (pink shading) (see **c** for cell numbers in each sample). **c**, Summary of clinical, pathological and GoT data from patients with *CALR*-mutated myeloproliferative neoplasms. BM, bone marrow; PB, peripheral blood. **d**, Number of genes per cell (left) and number of UMIs per cell (right) from published standard 10x data of healthy control CD34$^+$ cells and 10x data from 3′ v.2 chemistry of CD34$^+$ cells from patient samples that underwent concurrent GoT, after random downsampling of the reads from each sample to 50 million reads × 3 iterations, showing that the extra cycle of PCR and portioning a small aliquot from the 10x cDNA library for GoT using 3′ v.2 chemistry does not compromise scRNA-seq data.



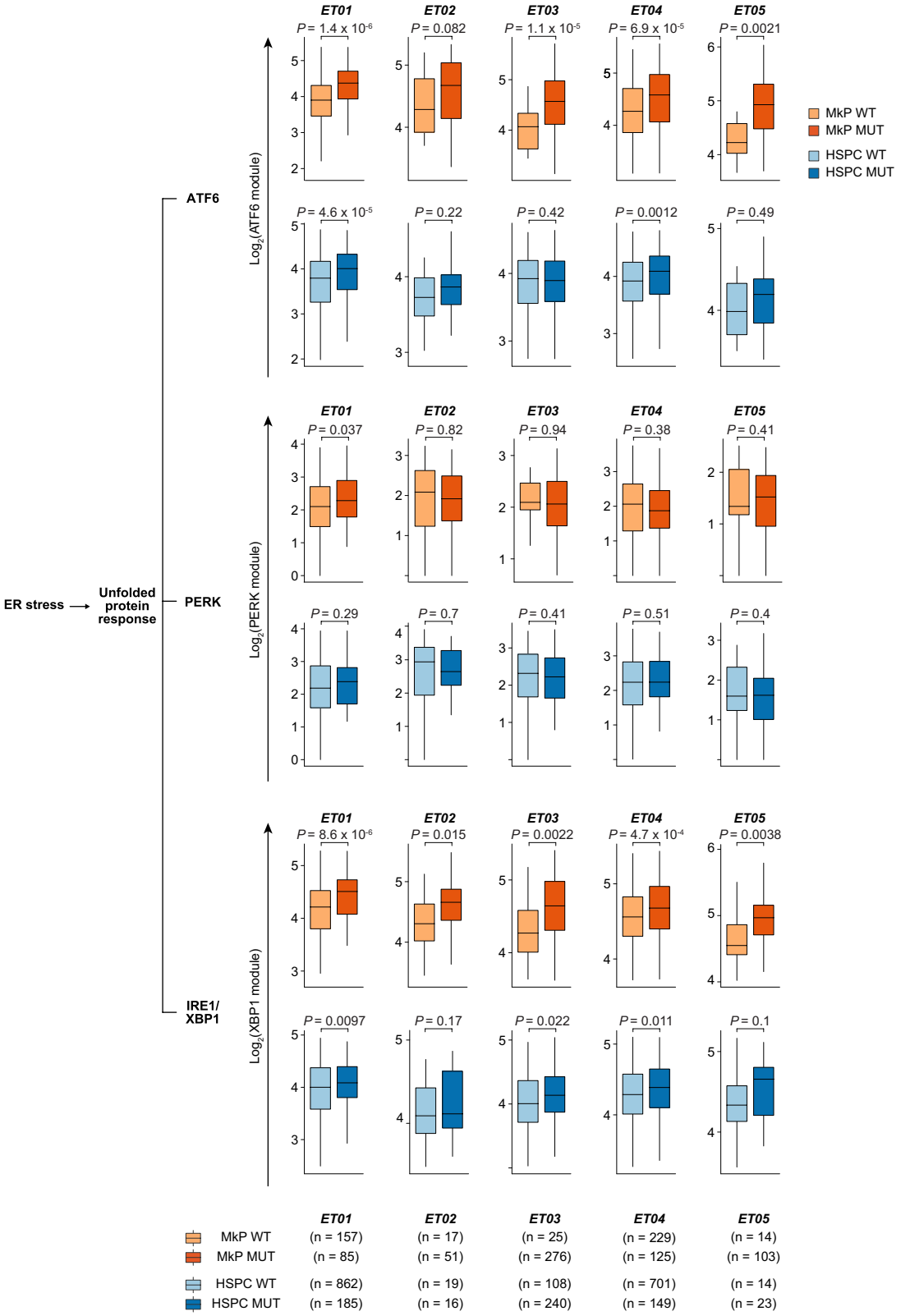
Extended Data Fig. 4 | Integration of samples from patients with essential thrombocythaemia and assignment of progenitor subsets. **a**, *t*-SNE projection of CD34$^+$ progenitor cells from samples ET01–ET05, after integration and batch correction using the Seurat package (Methods). **b**, Heat map of top ten differentially expressed genes for clusters; lineage-specific genes from a previous publication[26] are highlighted (Methods). **c**, Representative lineage-specific genes projected onto the *t*-SNE representation of CD34$^+$ cells from samples from patients with essential thrombocythaemia. **d**, *t*-SNE projection of CD34$^+$ cells from samples ET01–ET05 after applying a deep generative modelling approach for the single-cell analysis using the scVI package (Methods)[19], showing assignments of progenitor subsets as determined after clustering the cells using the Seurat package. **e**, Genotyping data from GoT are projected onto the *t*-SNE representation generated after the scVI analysis of progenitor cells from samples ET01–ET05. Cells without any GoT data are labelled NA (not assignable).



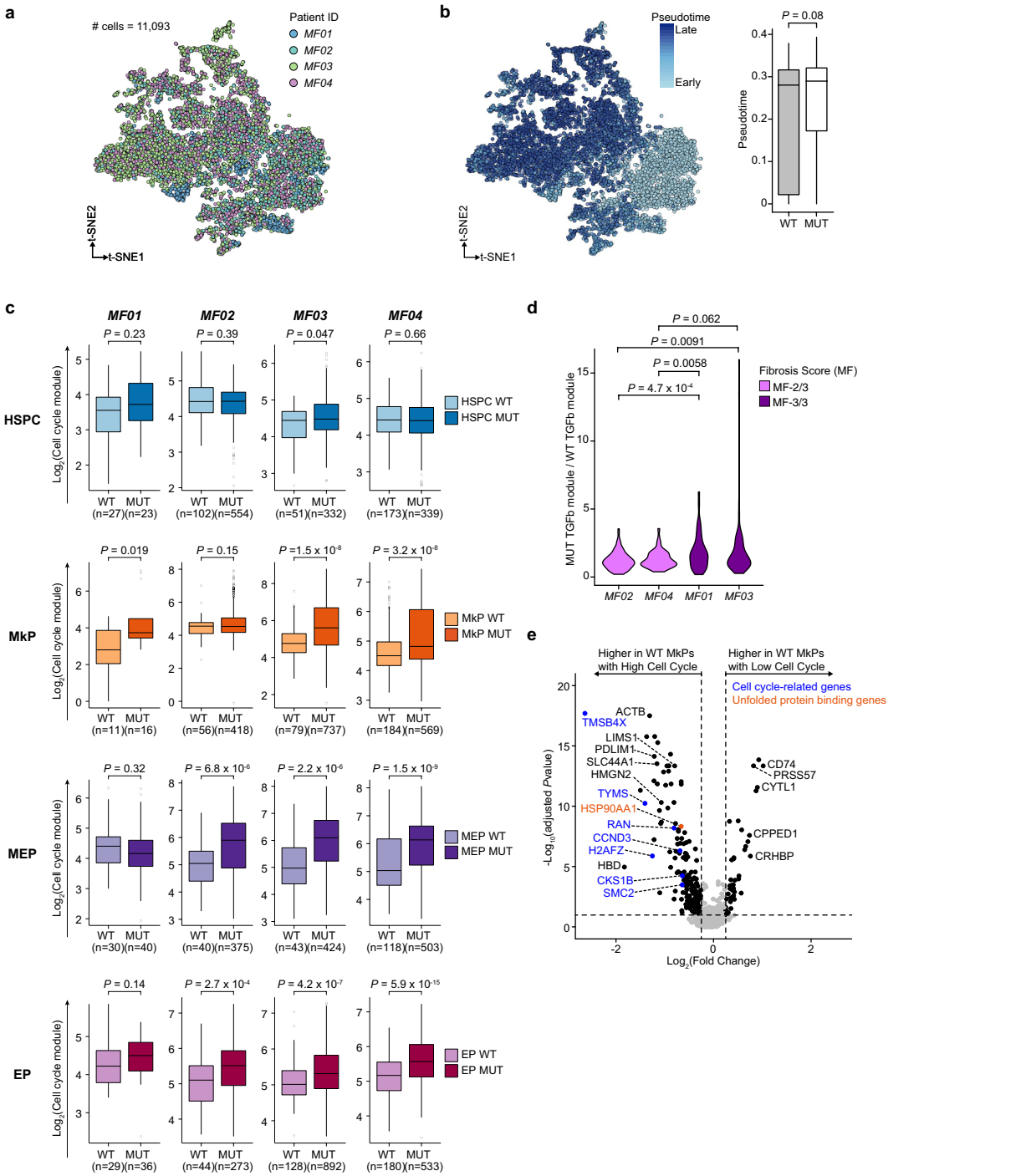
Extended Data Fig. 5 | Results of GoT analysis are robust to various amplicon UMI thresholds and linear modelling. a, Frequency of wild-type and mutant cells in HSPCs and MkPs with variable minimum genotyping UMI thresholds (two-sided Fisher's exact test; see Supplementary Table 6 for sample size). **b**, Pseudotime comparison between wild-type and mutant cells with an increasing number of thresholds for targeted genotyping UMI (two-sided *t*-test; see Supplementary Table 6 for sample size). **c**, Pseudotime comparison between mutant and wild-type cells with UMI threshold of 1 (same datasets as **b**), with statistical test using a generalized linear model including mutation status and total number of amplicon UMIs per cell. **d**, Across 100 iterations, the genotyping amplicon UMIs were downsampled to one per cell and the mutant-cell frequency was determined for MkPs or precursor B cells. This frequency was then divided by the total mutant-cell frequency across all progenitor subsets for each of the 100 iterations. Mean ± s.d. after $n = 100$ downsampling iterations (two-sided Wilcoxon rank-sum test). Essential thrombocythaemia samples with at least 20 cells in each cluster were analysed. **e**, VAF of *CALR* mutation in $CD34^+CD38^-$ (left), $CD34^+CD38^+$ (middle) and $CD34^+CD10^+$ (right) FACS-sorted peripheral blood cells from patients with essential thrombocythaemia determined by ddPCR.



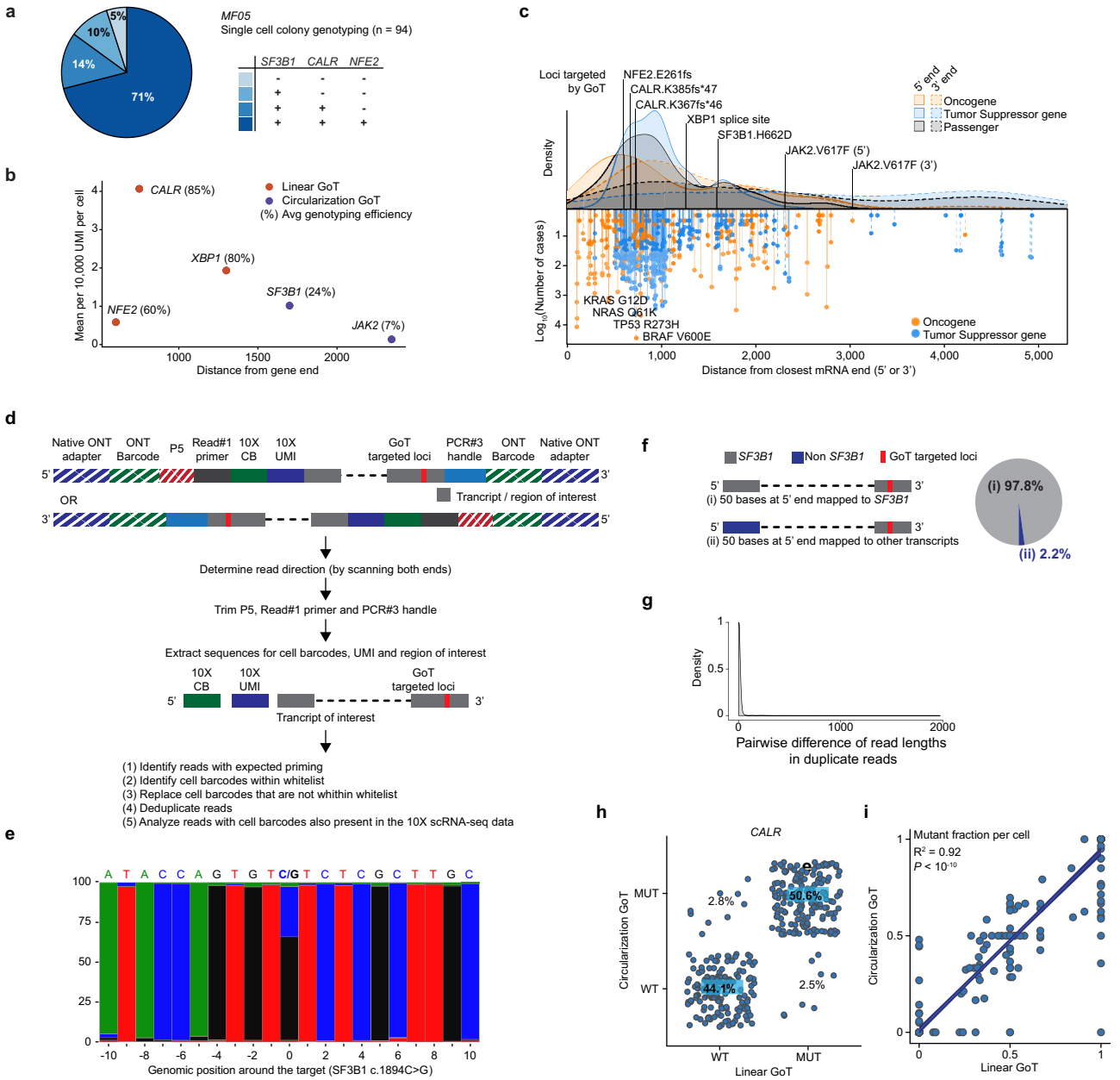
Extended Data Fig. 6 | Cell-cycle module expression in mutant and wild-type progenitor cells. a, S-phase and G2- and M-phase gene-module expression in wild-type and mutant cells, in HSPC and MkP clusters from essential thrombocythaemia samples. Cell-cycle module score represents the sum of S-phase and G2- and M-phase gene-module expression (two-sided Wilcoxon rank-sum test; see Methods and Supplementary Table 6 for sample size). Analysis was performed for clusters with at least 20 cells. **b**, Ratio of committed-progenitor priming-module expression of mutant and wild-type HSPCs. One mutant and one wild-type HSPC were randomly sampled from samples ET01–ET05 for each round of analysis ($n = 1{,}000$ iterations, two-sided Wilcoxon rank-sum test).



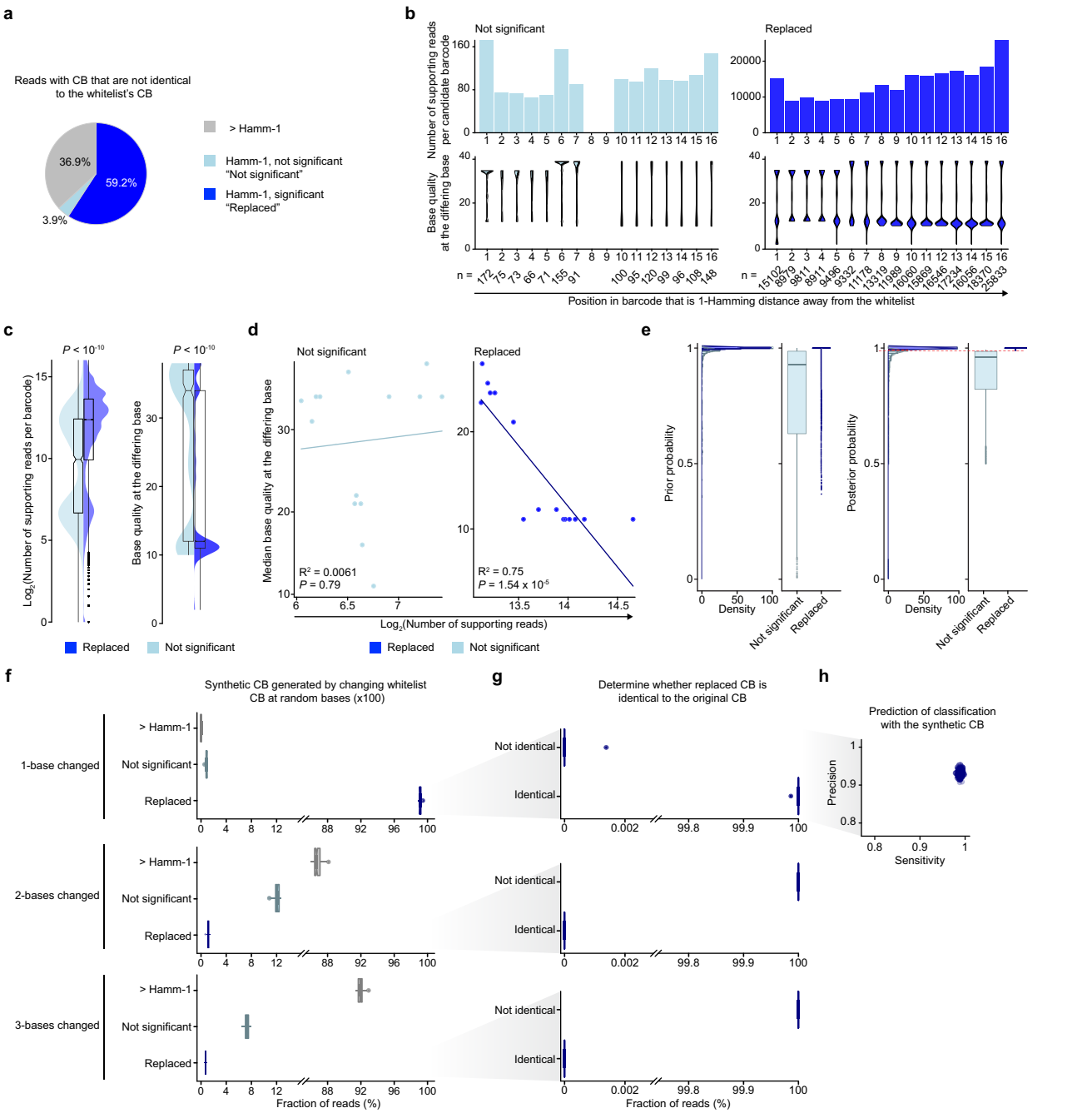
Extended Data Fig. 7 | ATF6 and IRE1 branches of the UPR are activated in *CALR*-mutated progenitor cells. By sample (ET01–ET05), expression of ATF6-, PERK- and XBP1-target genes in the UPR in *CALR* wild-type and mutant MkPs and HSPCs (two-sided Wilcoxon rank-sum test).



Extended Data Fig. 8 | *CALR*-mutated haematopoietic progenitor cells from myelofibrosis show upregulation of the IRE1-mediated UPR. **a**, *t*-SNE projection of $CD34^+$ progenitor cells from samples MF01–MF04, after integration and batch correction using the Seurat package (Methods) ($n = 11{,}093$). **b**, Left, *t*-SNE projection of $CD34^+$ progenitor cells from samples MF01–MF04 labelled with pseudotime[21] ($n = 11{,}093$). Right, pseudotime comparison between wild-type ($n = 2{,}221$) and mutant ($n = 7{,}483$) cells. *P* values from likelihood ratio tests of linear mixed model with genotype as fixed effect and individual patient samples as random effect, against the model without the genotype effect (Methods). **c**, Cell-cycle module score comparison between wild-type and mutant cells in patients with myelofibrosis (two-sided Wilcoxon rank-sum test). **d**, Ratio of TGFβ-signalling-pathway gene expression of mutant and wild-type MkPs. One mutant and one wild-type MkP were randomly sampled for each round of analysis ($n = 100$ iterations; two-sided Wilcoxon-rank sum test). **e**, Differentially expressed genes between wild-type MkPs with high cell-cycle expression ($n = 220$) and wild-type MkPs with low cell-cycle expression ($n = 110$), common across samples MF02–MF04. *P* values were combined using Fisher's combined test with Benjamini–Hochberg adjustments. Weighted average of fold change (expressed in $\log_2$) based on cell number across samples is shown (Methods).



Extended Data Fig. 9 | Deciphering subclonal progenitor identities using multiplex GoT, and targeting loci that are distant from transcript ends using circularization GoT. a, Single-cell cloning assay of peripheral blood cells from patient MF05 (Methods). **b**, Rate of targeted locus capture (per cent) as a function of gene expression and the distance of the targeted locus from the transcript ends. **c**, Distance of the mutation locus from transcript ends for pan-cancer drivers, and their frequencies (based on the number of times they are reported in the COSMIC database). Mutations are annotated as oncogenes, tumour-suppressor genes or passengers (as previously defined[60,61]). Relative density of each subclass of mutations from the closer end (that is, 3′ or 5′) is shown in the top panel. **d**, Schematic of analysis of ONT sequencing reads. **e**, Frequency of *SF3B1*-mutant and wild-type reads of linear GoT amplicon library sequenced with ONT. **f**, Analysis of *SF3B1* amplicon reads sequenced by ONT for inter-transcript PCR recombination by mapping 50 bp at the opposite end of the targeted locus, showing only the 2.2% of fragments that reflect inter-transcript recombination. **g**, Pairwise difference of read lengths for duplicate reads (that is, reads with the same cell barcode and UMI) of the *SF3B1* amplicon library sequenced with ONT, showing consistent read lengths of duplicate reads that support a low rate of intra-transcript PCR recombination. **h**, Comparison of genotype assignment for *CALR* in sample MF01 between linear GoT and circularization GoT after downsampling reads to 300,000 with 10 iterations ($n = 320$ cells). **i**, Comparison of *CALR*-mutant UMI fraction per cell in sample MF01 between linear GoT and circularization GoT after downsampling reads to 300,000 with 10 iterations ($n = 320$ cells, Pearson's correlation, *F*-test).



Extended Data Fig. 10 | Evaluation of barcode replacement in IronThrone GoT processing. a, Fraction of reads with cell barcodes that are not perfectly matched to the whitelisted cell barcodes from the species-mixing experiment. '>Hamm-1' denotes filtered reads with barcodes that are more than one Hamming distance away from whitelisted barcodes ($n = 139{,}422$ reads). 'Not significant' denotes filtered reads with barcodes that are one Hamming distance away from the whitelisted barcodes, but which have a low probability of originating from the barcode (posterior probability < 0.99, $n = 14{,}830$ reads). 'Replaced' denotes rescued reads with barcodes that have candidates that are one Hamming distance away from the whitelisted barcodes, with statistical significance (posterior probability ≥ 0.99, $n = 224{,}085$ reads). **b**, **c**, Number of supporting reads per candidate barcode and base quality at the differing base positions (**b**) and across base positions (**c**). Two-sided Wilcoxon rank-sum tests were applied to compare not significant ($n = 14{,}830$) and replaced ($n = 224{,}085$) barcodes. **d**, Correlation between the number of supporting reads per candidate barcode and median base quality at the differing base (two-tailed Pearson's correlation, *F*-test). **e**, Distribution of prior and posterior probabilities from not significant ($n = 14{,}830$) and replaced ($n = 224{,}085$) barcodes. The dashed red line represents the posterior probability cut-off (0.99). **f**–**h**, To further evaluate the efficiency of barcode replacement, we generated synthetic cell barcodes by randomly changing one base in whitelisted cell barcodes ($n = 100$ iterations). **f**, Percentage of reads with cell barcodes that are not identical to the whitelisted cell barcodes ($n = 100$ iterations). Percentages of replaced reads were 99.1% ± 0.001% (median ± absolute deviation) in simulations with 1 base changed, 1.1% ± 0.002% in simulations with 2 bases changed and 0.7 ± 0.001% in simulations with 3 bases changed. **g**, Determination of whether replaced cell barcodes are identical to the original cell barcodes. In simulations with 1 base change, the percentage of reads with replaced cell barcodes that were identical to the original cell barcodes was 100 ± 0% (median ± absolute deviation of 100 iterations). **h**, Estimation of prediction power for classifying cell barcodes from simulations with 1 base changed ($n = 100$ iterations).

nature research

Corresponding author(s): Dan Landau

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No commercial, open source, or custom software or code was used for data collection.
Data analysis	Conventional 10x data was processed using Cell Ranger (ver.2.1.0). Amplicon data was processed using the in-house codes which are available upon request. Downstream analyses and statistical analyses were performed using Seurat (ver. 3.0), Monocle (ver. 2.8), URD (ver.1.0.2), randomForest (ver. 4.6-14), gProfileR (ver. 0.6.7), lme4 package (ver. 1.2-1), R Stats (v3.5.1) packages under R environment (ver. 3.4.3) or scVI under Python environment (ver. 3.7.2). ONT data were basecalled by ONT Guppy (ver. 2.3.1), adaptor sequences trimmed using Porechop, and aligned with minimap2 (ver. 2.16). The Cancer Genome Atlas (TCGA) was analyzed using SAMtools mpileup (ver. 1.9) and biomaRt R package (ver. 2.39.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the raw data used in this study has been deposited at Gene Expression Omnibus, under the accession number GSE117826

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	A total of 29,815 cells were included from ET (n = 5) and MF (n = 4) patients. This enabled comparison of mutant and wildtype cells within each patient sample within the context of cell identity; data from replicate patient samples confirmed the findings within each patient.
Data exclusions	Based on the conventional practices of the scRNA-seq field set by the experts, we filtered out poor quality cells in 10x data analysis with following criteria: mitochondrial RNA fraction (>10%), minimal detected number of genes (<200), maximal detected number of genes (>mean +3*standard deviation across cells). In the amplicon data analysis, we applied specific thresholds (mismatch ratio <0.2, total number of duplicates >=2) based our experimental data (species-mixing study). Detailed methods for processing amplicon data are fully described in the online methods section and pipeline available on GitHub at https://github.com/thinktank-Q/IronThrone-GoT.
Replication	To verify the analyzed results, we have included five CALR+ ET and four CALR+ MF patient samples. The main findings of the study were reproduced in these replicate samples
Randomization	Since the analyses are based on comparison between mutant and wildtype cells within each patient samples, randomization was not relevant for this study.
Blinding	Since the analyses are based on comparison between mutant and wildtype cells within each patient samples, blinding was not relevant for this study.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CD34-PE-Vio770 (clone AC136, lot #5180718070, Miltenyi Biotec) for GoT; CD34-PeCy7 (clone 561, lot #B257238, BioLegend), CD38-APC (clone HIT2, lot #B247250, BioLegend) and CD10-FITC (clone HI10a, lot#B254556, BioLegend) for ddPCR. The antibodies were diluted according to manufacturers' recommendations.
Validation	Negative control has been performed on every patient sample to properly set-up the gate of CD34, CD38, and CD10 positive cell

populations. For CD34-PE-Vio770, CD34 mRNA expression has been further observed in the sorted cells by scRNA-seq. CD34, CD38 and CD10 antibodies used for ddPCR passed the quality control testing for flow cytometric analysis by the manufacturer.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	TF1 and HEL cell lines were obtained from ATCC; UT7-MPL-CALR WT and Ba/F3-MPL-CALR MUT cells were provided by Elf, et al (Cancer Discovery, 2016).
Authentication	Both TF1 and HEL cell lines have been only used as sources of cells with JAK2 WT and JAK2 V617F, respectively, which we have confirmed previously by sequencing the genomic DNA of those cell lines. UT7-MPL-CALR WT and Ba/F3-MPL-CALR MUT cells were validated by Elf, et al (Cancer Discovery, 2016) and also confirmed in our GoT data.
Mycoplasma contamination	The cell lines have been tested negatively for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	TF1 and HEL cell lines were used for the sole purpose of extracting RNA for JAK2 WT or JAK2 V617F. UT7-MPL-CALR WT and Ba/F3-MPL-CALR MUT cells were used for the sole purpose of validating GoT procedure for identifying CALR mutant and wildtype status.

Human research participants

Policy information about studies involving human research participants

Population characteristics	Study included cryopreserved bone marrow or peripheral blood samples from patients with documented CALR-mutated or JAK2 myeloid neoplasms (6 with essential thrombocythemia, 5 with myelofibrosis, age range: 17-69, gender: 5F, 6M) which had been banked under specimen acquisition and molecular profiling protocols approved by the Institutional Review Board of Memorial Sloan-Kettering Cancer Center or Weill Cornell Medicine (Supplementary Table 1). None of the patients had been received chemotherapy at the time of sample collection. All patients provided informed consent.
Recruitment	Samples were obtained from patients who sought consultation/were being treated at Memorial Sloan-Kettering Cancer Center or Weill Cornell Medicine.

ARTICLE

https://doi.org/10.1038/s41586-019-1346-5

High-dimensional geometry of population responses in visual cortex

Carsen Stringer[1,2,6]*, Marius Pachitariu[1,3,6]*, Nicholas Steinmetz[3,5], Matteo Carandini[4,7] & Kenneth D. Harris[3,7]*

A neuronal population encodes information most efficiently when its stimulus responses are high-dimensional and uncorrelated, and most robustly when they are lower-dimensional and correlated. Here we analysed the dimensionality of the encoding of natural images by large populations of neurons in the visual cortex of awake mice. The evoked population activity was high-dimensional, and correlations obeyed an unexpected power law: the *n*th principal component variance scaled as $1/n$. This scaling was not inherited from the power law spectrum of natural images, because it persisted after stimulus whitening. We proved mathematically that if the variance spectrum was to decay more slowly then the population code could not be smooth, allowing small changes in input to dominate population activity. The theory also predicts larger power-law exponents for lower-dimensional stimulus ensembles, which we validated experimentally. These results suggest that coding smoothness may represent a fundamental constraint that determines correlations in neural population codes.

The visual cortex contains millions of neurons, and the patterns of activity that images evoke in these neurons form a 'population code'. The structure of this code is largely unknown, due to the lack of techniques that are able to record from large populations. Nonetheless, the population code is the subject of long-standing theories.

One such theory is the efficient coding hypothesis[1–3], which suggests that the neural code maximizes the transmission of information by eliminating correlations in natural image inputs. Such codes are high-dimensional and sparse, which can enable complex features to be read out by simple downstream networks[4–6].

However, several studies have suggested that neural codes are confined to low-dimensional subspaces (or 'planes')[7–15]. Codes of low planar dimension are correlated and redundant, allowing for robust computations of stimuli despite the presence of noise[16,17]. Nevertheless, low planar dimension is inevitable given stimuli or tasks of limited complexity[18]: the responses to a set of n stimuli, for example, have to lie in an n-dimensional subspace. The planar dimension of the cortical code thus remains an open question, which can only be answered by recording the responses of large numbers of neurons to large numbers of stimuli.

Here we recorded the simultaneous activity of approximately 10,000 neurons in the mouse visual cortex, in response to thousands of natural images. We found that stimulus responses were neither uncorrelated ('efficient coding') nor low-dimensional. Instead, responses occupied a multidimensional space, with the variance in the nth dimension scaling as a power law $n^{-\alpha}$, where $\alpha \approx 1$. We showed mathematically that if variances decay more slowly than a power law with exponent $\alpha = 1 + 2/d$, where d is the dimension of the input ensemble, then the space of neural activity must be non-differentiable—that is, not smooth. We varied the dimensionality of the stimuli d and found that the neural responses respected this lower bound. These findings suggest that the population responses are constrained by efficiency, to make best use of limited numbers of neurons, and smoothness, which enables similar images to evoke similar responses.

Simultaneous recordings of over 10,000 neurons

To obtain simultaneous recordings of approximately 10,000 cells from mouse V1, we used resonance-scanning two-photon calcium microscopy, using 11 imaging planes spaced at 35 μm (Fig. 1a). The slow time course of the GCaMP6s sensor enabled activity to be detected at a scan rate of 2.5 Hz, and an efficient data processing pipeline[19] yielded the activity of a large numbers of cells (Fig. 1b). Natural image scenes obtained from the ImageNet database[20] were presented on an array of three monitors surrounding the mouse (Fig. 1c), at an average of one image per second. Cells were tuned to these natural image stimuli: in experiments in which responses to 32 images were averaged over 96 repeats (Fig. 1d), stimulus responses accounted for 55.4 ± 3.3% (mean ± s.e.m., $n = 4$ recordings) of the trial-averaged variance. Consistent with previous reports[21–23], neuronal responses were sparse: only a small fraction of cells (13.4 ± 1.0%; mean ± s.e.m., $n = 4$ recordings) were driven more than two standard deviations above their baseline firing rate by any particular stimulus.

For our main experiments we assembled a sequence of 2,800 image stimuli. These stimuli were presented twice in the same order, to maximize the number of images presented while still allowing analyses based on cross-validation (Fig. 1e). Most neurons (81.4 ± 5.1%; mean ± s.e.m., $n = 7$ recordings) showed correlation between repeats at $P < 0.05$ (Extended Data Fig. 1a, b). Nevertheless, consistent with previous reports[24], the responses showed substantial trial-to-trial variability. Cross-validation showed that stimulus responses accounted for, on average, 13.2 ± 1.5% of the single-trial variance (Extended Data Fig. 1c), and the average signal-to-noise ratio was 17.3 ± 2.4% (Fig. 1f). This level of trial-to-trial variability was not due to our particular recording method: measuring responses to the same stimuli electrophysiologically yielded a similar signal-to-noise ratio (Extended Data Fig. 2). Despite this trial-to-trial variability, however, population activity recorded during a single trial contained substantial information about the sensory stimuli. A simple nearest-neighbour decoder, trained on one repeat and tested on the other, was able to identify the presented stimulus with up to 75.5% accuracy (Fig. 1g; range 25.4–75.5%; median

[1]HHMI Janelia Research Campus, Ashburn, VA, USA. [2]UCL Gatsby Computational Neuroscience Unit, University College London, London, UK. [3]UCL Institute of Neurology, University College London, London, UK. [4]UCL Institute of Ophthalmology, University College London, London, UK. [5]Present address: Department of Biological Structure, University of Washington, Seattle, WA, USA. [6]These authors contributed equally: Carsen Stringer, Marius Pachitariu. [7]These authors jointly supervised this work: Matteo Carandini, Kenneth D. Harris. *e-mail: stringerc@janelia.hhmi.org; pachitariu@janelia.hhmi.org; kenneth.harris@ucl.ac.uk

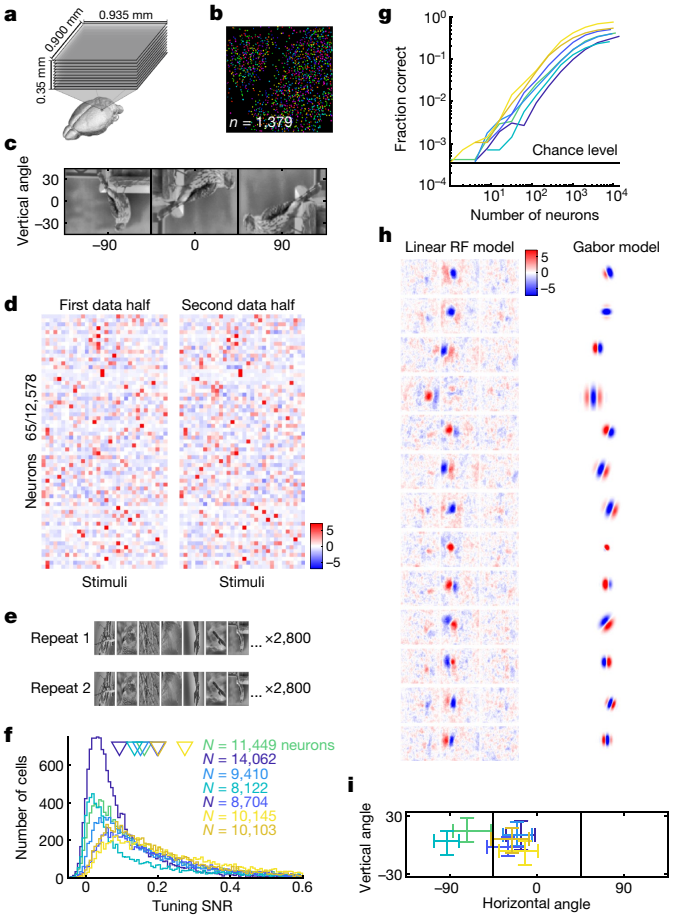


Fig. 1 | Population coding of visual stimuli. a, Simultaneous recording of approximately 10,000 neurons using 11-plane two-photon calcium imaging. **b**, Randomly pseudocoloured cells in an example imaging plane. **c**, An example stimulus spans three screens surrounding the head of the mouse. **d**, Mean responses (trial-averaged) of 65 randomly chosen neurons to 32 image stimuli (96 repeats, *z*-scored, scale bar represents standard deviations, one recording out of four is shown). **e**, A sequence of 2,800 stimuli was repeated twice during the recording. **f**, Neural stimulus tuning. The plot shows the distribution of single-cell signal-to-noise ratios (SNR) (2,800 stimuli, two repeats). Colours denote recordings; arrows represent means. **g**, Stimulus decoding accuracy as a function of neuron count for each recording. **h**, Example receptive fields (RFs) fit using reduced-rank regression or Gabor models (*z*-scored) (one recording shown, out of seven). **i**, Distribution of the receptive field centres, plotted on the left and centre screens (lines denote screen boundaries). Each cross represents a different recording, with 95% of the receptive field centres of the neurons within the error bars.

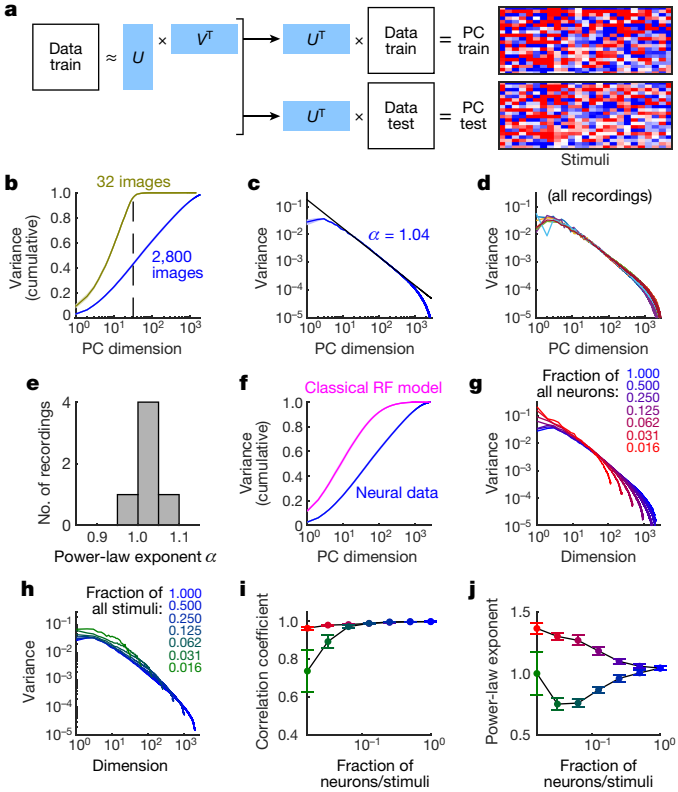


Fig. 2 | Visual cortical responses are high-dimensional with power-law eigenspectra. a, The eigenspectrum of visual stimulus responses was estimated by cvPCA, projecting singular vectors from the first repeat onto responses from the second. PC, principal component. **b**, Cumulative fraction of variance in planes of increasing dimension, for an ensemble of 2,800 stimuli (blue) and for 96 repeats of 32 stimuli (green). The dashed line indicates 32 dimensions. **c**, Eigenspectrum plotted in descending order of training-set singular value for each dimension, averaged across 7 recordings (shaded error bars represent s.e.m.). The black line denotes the linear fit of $1/n^{\alpha}$. **d**, Eigenspectra of each recording plotted individually. **e**, Histogram of power-law exponents α across all recordings. **f**, Cumulative eigenspectrum for a simple/complex Gabor model fit to the data (pink) superimposed on the true data (blue). **g**, Eigenspectra computed from random subsets of recorded neurons. Different colours indicate the different fractions of neurons. **h**, The same analysis as in **g**, but for random subsets of stimuli. **i**, Pearson correlation of log variance and log dimension over dimensions 11–500, as a function of fraction analysed (1 indicates a power law). **j**, Power-law exponents of the spectra plotted in **g**, **h**.

41.7% compared to a chance level of 0.036%, $n = 7$ recordings). The decoding accuracy did not saturate at a population size of 10,000, which suggests that performance would further improve with even larger neural populations.

The visual properties of neurons were consistent with those reported previously[23,25], and were highly diverse across the population. The responses of the neurons were only partially captured by classical linear–nonlinear models, which is consistent with previous studies of the visual cortex[26–30]. We calculated a receptive field for each cell from its responses to natural images in two ways: by fitting linear receptive fields regularized with a reduced-rank method; or by searching for an optimal Gabor filter that was rectified to simulate simple cell responses, and quadrature filtered to simulate complex cell responses. As expected[26–30], both receptive field models explained only a minor portion of the stimulus-related variance: the linear model explained 11.4 ± 0.7% (mean ± s.e.m.), and the Gabor model explained 18.5 ± 1.0% (mean ± s.e.m., $n = 7$ recordings each). As expected from retinotopy, there was overlap between the receptive field locations of simultaneously recorded neurons, but the sizes and shapes of the receptive fields were highly diverse (Fig. 1h, Extended Data Fig. 3).

Power–law scaling of dimensionality

To characterize the geometry of the population code for visual stimuli, we developed a method of cross-validated principal component analysis (cvPCA). cvPCA measures the reliable variance of stimulus-related dimensions, excluding trial-to-trial variability from unrelated cognitive and/or behavioural variables or noise. It accomplishes this by computing the covariance of responses between training and test presentations of an identical stimulus ensemble (Fig. 2a). Because only stimulus-related activity will be correlated across presentations, cvPCA provides an unbiased estimate of the stimulus-related variance. In simulations that use the same noise statistics as our recordings, we confirmed that this technique recovers the true variances (Extended Data Figs. 4, 5, Supplementary Discussion 1).

This method revealed that the visual population responses did not lie on any low-dimensional plane within the space of possible firing patterns. The amount of variance explained continued to increase as

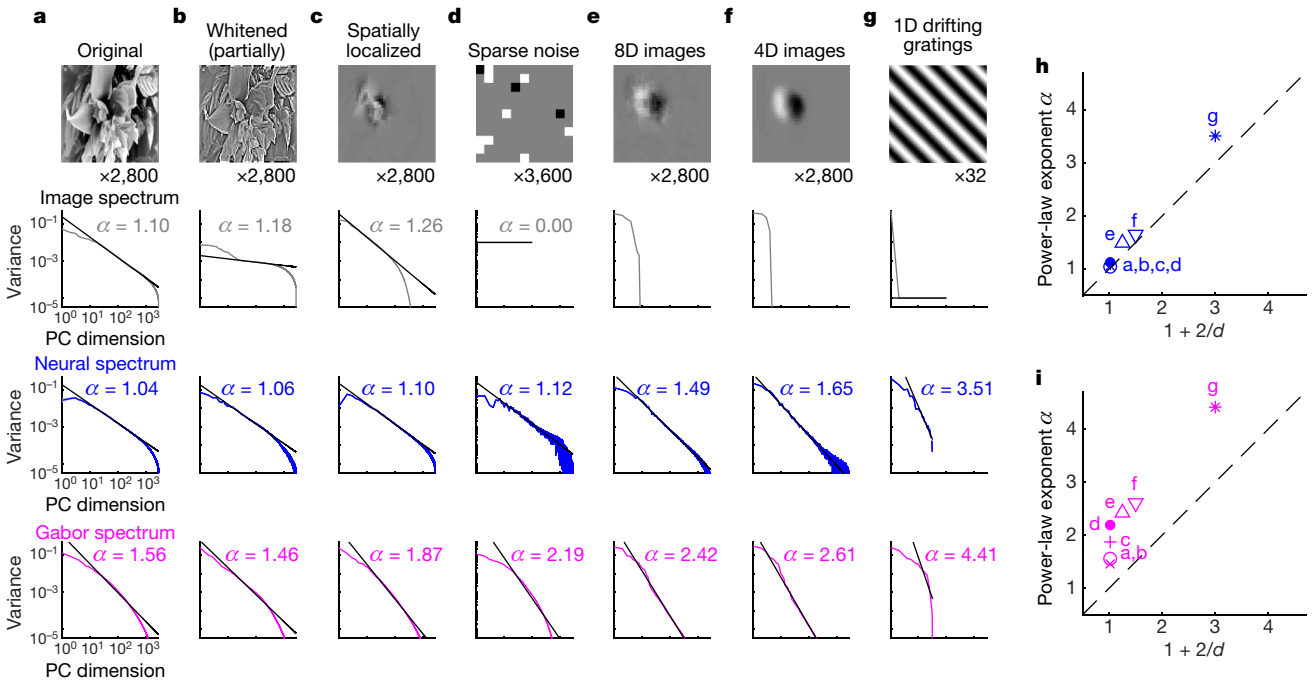


Fig. 3 | Power-law exponent depends on input dimensionality, but not on image statistics. a–**g**, Examples of presented images (top), eigenspectra of image pixel intensities (second row, grey), eigenspectra of visual cortical responses (third row, blue), and eigenspectra of responses of the Gabor receptive field model (fourth row, pink), for the original images (**a**), spatially whitened images lacking $1/n$ image spectrum (**b**), images windowed over the receptive field of the recorded population (**c**), sparse noise stimuli (**d**), images projected into eight dimensions, which produces a faster neural eigenspectrum decay with exponent $\alpha = 1.49$ (**e**), images projected into 4 dimensions, for which $\alpha = 1.65$ (**f**), drifting gratings, a one dimensional stimulus ensemble, for which $\alpha = 3.51$ (**g**). **h**, **i**, Summary of power-law exponents α for neural responses (**h**) and the Gabor model (**i**), as a function of the dimensionality of the stimulus set d. The dashed line, at $\alpha = 1 + 2/d$, corresponds to the border of fractality.

further dimensions were included, without saturating at any dimensionality below the maximum possible (Fig. 2b). As a control analysis, we applied cvPCA to the neural responses obtained when only 32 images were shown many times— the reliable component of these responses must, by definition, lie in a 32-dimensional subspace—and as expected we observed a saturation of the variance after 32 dimensions.

The analysis of our data using cvPCA revealed an unexpected finding: the fraction of neural variance in planes of successively larger dimensions followed a power law. The eigenspectrum—the function summarizing the variance of the nth principal component—had a magnitude that was approximately proportional to $1/n$ (Fig. 2c); this reflects successively less variance in dimensions that encode finer stimulus features (Extended Data Fig. 6). The power-law structure did not result from averaging over experiments: analysis of data from each mouse individually revealed power-law behaviour in every case (Fig. 2d). The scaling exponent of the power law was on average just above 1 (1.04 ± 0.02; mean ± s.e.m., $n = 7$ recordings, Fig. 2e). This eigenspectrum reflected correlations between neurons, and was not the consequence of a log-normal distribution of firing rates or signal variance (Extended Data Fig. 7). In addition, this result could not be explained by classical models of visual cortical receptive fields: the model of visual responses based on Gabor receptive fields with parameters fit to single cell responses (Fig. 1h) had lower dimensionality than the neural responses (Fig. 2f).

The range of dimensions over which the power law held grew with the number of neurons and stimuli that were analysed. To show this, we repeated the analyses on randomly chosen subsets of neurons or stimuli (Fig. 2g, h). Both the correlation coefficient and the slope (which represents the power-law exponent) approached 1 for increasing subset sizes (Fig. 2i, j, Extended Data Fig. 8). Electrophysiological recordings—with fewer recorded neurons and fewer presented stimuli—had the same eigenspectrum as a similarly-sized subset of the two-photon data (Extended Data Fig. 9). We conclude that the power law held accurately over approximately two orders of magnitude in these recordings, and we infer that it would probably extend further if more neurons and stimuli could be analysed.

Power-law and stimulus statistics

The power law followed by the neural eigenspectrum could not be explained by the well-known power-law structure of natural images[31,32] (Fig. 3a). To show this, we removed the image power law by spatially whitening the images, and presented the whitened stimuli to three of the mice. Although the power law in the image pixels was abolished, the power law in the neural responses remained (Fig. 3b). Furthermore, the eigenspectrum of neural responses could not be explained by straightforward receptive field properties: the model of visual responses based on Gabor receptive fields produced eigenspectra that decayed more quickly than the actual responses, and were worse fit by a power law ($P < 10^{-3}$, Wilcoxon rank-sum test on Pearson correlations, Fig. 3a, b).

The power-law eigenspectra also did not arise from other characteristics of natural images. To investigate the role of long-range image correlations, we constructed spatially localized image stimuli, in which the region outside the classical receptive field was replaced by grey. Again, the power law persisted with an exponent close to 1 (Fig. 3c). Finally, we removed any natural image structure and recorded responses to sparse noise stimuli (Fig. 3d). Again, we observed a power-law spectrum with an exponent close to 1 (1.13 ± 0.04; mean ± s.e.m., $n = 3$ recordings); although it was higher than for the natural image stimuli ($P = 0.067$, Wilcoxon two-sided rank-sum test). As with natural images, these power laws became more accurate the more neurons and stimuli were analysed (Extended Data Fig. 10). We therefore conclude that the power-law spectra exhibited by neural populations do not reflect the neural processing of a property that is specific to natural images.

Power-law and stimulus dimensionality

Power-law eigenspectra are observed in many scientific domains, and are related to the smoothness of the underlying functions. For example, if a function of one variable is differentiable, its Fourier spectrum must

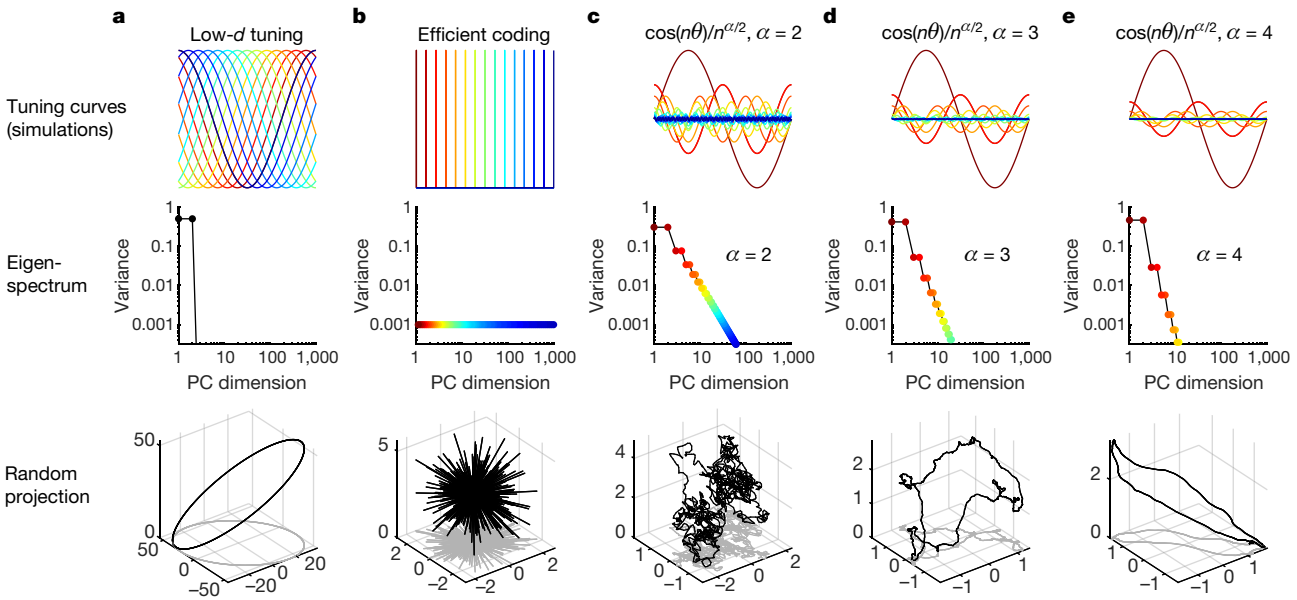


Fig. 4 | The smoothness of simulated neural activity depends on the eigenspectrum decay. Simulations of neuronal population responses to a one-dimensional stimulus (horizontal axis) (top), their eigenspectra (middle), and a random projection of population responses into three-dimensional space (bottom). **a**, Wide tuning curves, corresponding to a circular neural manifold in a two-dimensional plane. **b**, Narrow tuning curves corresponding to uncorrelated responses as predicted by the efficient coding hypothesis. **c**–**e**, Scale-free tuning curves corresponding to power-law variance spectra, with exponents of 2, 3 (the critical value for $d = 1$) or 4. The tuning curves in **c**–**e** represent PC dimensions rather than individual simulated neurons. Blue, dimensions that encode fine stimulus details; red, dimensions that encode large-scale stimulus differences.

decay asymptotically faster than a power law of exponent 1 (see, for example, a previously published study[33]). We therefore theorized that the variance power law might be related to smoothness of the neural responses. We showed mathematically that if the sensory stimuli presented can be characterized by d parameters, and if the mapping from these parameters to (noise-free) neural population responses is differentiable, then the population eigenspectrum must decay asymptotically faster than a power law of exponent $\alpha = 1 + 2/d$ (Supplementary Discussion 2). Conversely, if the eigenspectrum decays slower than this, a smooth neural code is impossible: its derivative tends to infinity with the number of neural dimensions, and the neural responses must lie on a fractal rather than a differentiable manifold.

This mathematical analysis gave rise to an experimental prediction. For a high-dimensional stimulus ensemble such as a set of natural images, d will be large and so $1 + 2/d \approx 1$, which is close to the exponent that we observed. However, for smaller values of d, the power law must have larger exponents if fractality is to be avoided. We therefore hypothesized that lower-dimensional stimulus sets would evoke population responses with larger power-law exponents. We obtained stimulus ensembles of dimensionality $d = 8$ and $d = 4$ by projecting the natural images onto a set of d basis functions (Fig. 3e, f). For a stimulus ensemble of dimensionality $d = 1$ we used drifting gratings, parameterized by a single direction variable. Consistent with the hypothesis, stimulus sets with $d = 8$, 4 and 1 yielded power-law scaling of eigenvalues with exponents of 1.49, 1.65 and 3.51, near but above the lower bounds of 1.25, 1.50 and 3.00 that are predicted by the $1 + 2/d$ exponent (Fig. 3h). The eigenspectra of simulated responses from a Gabor receptive field model fit to the data decayed even faster, suggesting a differentiable but lower-dimensional representation (Fig. 3i). These results suggest that the neural responses lie on a manifold that is almost as high-dimensional as is possible without becoming fractal.

Discussion

We found that the variance of the nth dimension of visual cortical population activity decays as a power of n, with exponent $\alpha \approx 1 + 2/d$ where d is the dimensionality of the space of sensory inputs. The population eigenspectrum reflects the fraction of neural variance that is devoted to representing coarse and fine stimulus features (Extended Data Fig. 6, Supplementary Discussion 2, 3). If the eigenspectrum were to decay slower than $n^{-1-2/d}$ then the neural code would emphasize fine stimulus features so strongly that it could not be differentiable. Our results therefore suggest that the eigenspectrum of the visual cortical code decays almost as slowly as is possible while still allowing smooth neural coding.

To illustrate the consequences of eigenspectrum decay for neural codes, we simulated various one-dimensional coding schemes in populations of 1,000 neurons, and visualized them by random projection into three-dimensional space (Fig. 4). The stimulus was parameterized by a single circular variable, such as the direction of a grating. A low-dimensional code with two non-zero eigenvalues produced a circular neural manifold (Fig. 4a). An uncorrelated, high-dimensional code in which each neuron responded to a different stimulus produced 1,000 equal variances, which is consistent with the efficient coding hypothesis (Fig. 4b). However this code did not respect distances: responses to stimuli separated by just a few degrees differed as much as responses to diametrically opposite stimuli, and the neural manifold appeared as a spiky, discontinuous ball. Power-law codes (Supplementary Discussion 2.7, example 2) show a scale-free geometry, the smoothness of which depends on the exponent α (Fig. 4c–e). A power-law code with $\alpha = 2$ (Fig. 4c) was a non-differentiable fractal because the many dimensions that encode fine stimulus details together outweighed the few dimensions that encode large-scale stimulus differences. At the critical exponent of $\alpha = 3$ (which is equal to $1 + 2/d$, because $d = 1$), the neural manifold was on the border of differentiability; the code represents fine differences between stimuli while still preserving large-scale stimulus features (Fig. 4d). A higher exponent led to a smoother neural manifold (Fig. 4e).

Neural representations with close-to-critical power-law eigenspectra may provide the brain with codes that are as efficient and flexible as possible while still allowing robust generalization. The efficient coding hypothesis suggests that information is optimally encoded when responses to different stimuli are as different as possible. However, such codes carry a cost: if the neural responses to any pair of stimuli were orthogonal, then stimuli that differ only in tiny details would have completely different representations (Supplementary Discussion 2.1). Similar behaviour can be seen in deep-neural-network architectures

that misclassify 'adversarial images' that differ imperceptibly from the training examples[34,35]. We suggest that a power-law code that is just above the critical exponent represents a balance between the efficiency of high-dimensional codes and the robustness of smooth codes, which enable generalization.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1346-5.

Received: 6 August 2018; Accepted: 29 May 2019;
Published online 26 June 2019.

1. Barlow, H. B. in *Sensory Communication* (ed. Rosenblith, W.) 217–234 (MIT Press, 1961).
2. Atick, J. J. & Redlich, A. N. Towards a theory of early visual processing. *Neural Comput.* **2**, 308–320 (1990).
3. Simoncelli, E. P. & Olshausen, B. A. Natural image statistics and neural representation. *Annu. Rev. Neurosci.* **24**, 1193–1216 (2001).
4. DiCarlo, J. J., Zoccolan, D. & Rust, N. C. How does the brain solve visual object recognition? *Neuron* **73**, 415–434 (2012).
5. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013).
6. Chung, S., Lee, D. D. & Sompolinsky, H. Classification and geometry of general perceptual manifolds. *Phys. Rev. X* **8**, 031003 (2018).
7. Cunningham, J. P. & Yu, B. M. Dimensionality reduction for large-scale neural recordings. *Nat. Neurosci.* **17**, 1500–1509 (2014).
8. Machens, C. K., Romo, R. & Brody, C. D. Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex. *J. Neurosci.* **30**, 350–360 (2010).
9. Kobak, D. et al. Demixed principal component analysis of neural population data. *eLife* **5**, e10989 (2016).
10. Archer, E. W., Koster, U., Pillow, J. W. & Macke, J. H. Low-dimensional models of neural population activity in sensory cortical circuits. In *Proc. 27th International Conference on Neural Information Processing Systems* (eds Ghahramani, Z. et al.) 343–351 (Curran, 2014).
11. Sadtler, P. T. et al. Neural constraints on learning. *Nature* **512**, 423–426 (2014).
12. Chapin, J. K. & Nicolelis, M. A. Principal component analysis of neuronal ensemble activity reveals multidimensional somatosensory representations. *J. Neurosci. Methods* **94**, 121–140 (1999).
13. Bathellier, B., Buhl, D. L., Accolla, R. & Carleton, A. Dynamic ensemble odor coding in the mammalian olfactory bulb: sensory information at different timescales. *Neuron* **57**, 586–598 (2008).
14. Churchland, M. M. et al. Neural population dynamics during reaching. *Nature* **487**, 51–56 (2012).
15. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013).
16. Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
17. Reich, D. S., Mechler, F. & Victor, J. D. Independent and redundant information in nearby cortical neurons. *Science* **294**, 2566–2568 (2001).
18. Gao, P. et al. A theory of multineuronal dimensionality, dynamics and measurement. Preprint at https://www.biorxiv.org/content/early/2017/11/12/214262 (2017).
19. Pachitariu, M. et al. Suite2p: beyond 10,000 neurons with standard two-photon microscopy. Preprint at https://www.biorxiv.org/content/early/2017/07/20/061507 (2016).
20. Deng, J. et al. Imagenet: A large-scale hierarchical image database. In *IEEE Conference on Computer Vision and Pattern Recognition* 248–255 (IEEE, 2009).
21. Vinje, W. E. & Gallant, J. L. Natural stimulation of the nonclassical receptive field increases information transmission efficiency in V1. *J. Neurosci.* **22**, 2904–2915 (2002).
22. Ringach, D. L. Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. *J. Neurophysiol.* **88**, 455–463 (2002).
23. Niell, C. M. & Stryker, M. P. Highly selective receptive fields in mouse visual cortex. *J. Neurosci.* **28**, 7520–7536 (2008).
24. Softky, W. R. & Koch, C. The highly irregular firing of cortical cells is inconsistent with temporal integration of random EPSPs. *J. Neurosci.* **13**, 334–350 (1993).
25. Cossell, L. et al. Functional organization of excitatory synaptic strength in primary visual cortex. *Nature* **518**, 399–403 (2015).
26. Smyth, D., Willmore, B., Baker, G. E., Thompson, I. D. & Tolhurst, D. J. The receptive-field organization of simple cells in primary visual cortex of ferrets under natural scene stimulation. *J. Neurosci.* **23**, 4746–4759 (2003).
27. Carandini, M. et al. Do we know what the early visual system does? *J. Neurosci.* **25**, 10577–10597 (2005).
28. David, S. V. & Gallant, J. L. Predicting neuronal responses during natural vision. *Netw. Comput. Neural Syst* **16**, 239–260 (2005).
29. Touryan, J., Felsen, G. & Dan, Y. Spatial structure of complex cell receptive fields measured with natural images. *Neuron* **45**, 781–791 (2005).
30. de Vries, S. E. J. et al. A large-scale, standardized physiological survey reveals higher order coding throughout the mouse visual cortex. Preprint at https://www.biorxiv.org/content/early/2018/06/29/359513 (2018).
31. Field, D. J. Relations between the statistics of natural images and the response properties of cortical cells. *J. Opt. Soc. Am. A* **4**, 2379–2394 (1987).
32. Ruderman, D. L. & Bialek, W. Statistics of natural images: Scaling in the woods. In *Advances in Neural Information Processing Systems* 551–558 (1994).
33. Tao, T. *An Epsilon of Room, I: Real Analysis* 1.12.3 (American Mathematical Society, 2010).
34. Szegedy, C. et al. Intriguing properties of neural networks. Preprint at http://arxiv.org/abs/1312.6199 (2013).
35. Goodfellow, I. J., Shlens, J. & Szegedy, C. Explaining and harnessing adversarial examples. Preprint at http://arxiv.org/abs/1412.6572 (2014).

Acknowledgements We thank M. Krumin for assistance with the two-photon microscopes, C. Reddy for surgeries, and K. Falconer, A. Gretton and M. Benna for discussions of mathematics. This research was funded by Wellcome Trust Investigator grants (108726, 205093 and 204915) and by a grant from the Simons Foundation (SCGB 325512). C.S. was funded by a four-year Gatsby Foundation PhD studentship. M.C. holds the GlaxoSmithKline/Fight for Sight Chair in Visual Neuroscience. K.D.H. was funded by the European Research Council (694401). N.S. was supported by postdoctoral fellowships from the Human Frontier Sciences Program and the Marie Curie Action of the EU (656528). C.S. and M.P. are now funded by HHMI Janelia.

Reviewer information *Nature* thanks Jakob Macke, Ken Miller and Byron Yu for their contribution to the peer review of this work.

Author contributions C.S., M.P., N.S., M.C. and K.D.H. conceptualized the study; C.S., M.P., N.S. and K.D.H. devised the methodology; C.S. and M.P. designed the software; C.S., M.P. and N.S. performed experiments; C.S. and M.P. analysed the data; K.D.H. proved mathematical theorems; C.S., M.P., N.S., M.C. and K.D.H. wrote the paper; and M.C. and K.D.H. provided resources and acquired funding.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1346-5.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1346-5.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to C.S., M.P. or K.D.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

All experimental procedures were conducted according to the UK Animals Scientific Procedures Act (1986). Experiments were performed at University College London under personal and project licenses released by the Home Office following appropriate ethics review.

Animals and surgery. We used mice that were bred to express GCaMP6s in excitatory neurons in our recordings: 13 recordings from TetO-GCaMP6s × Emx1-IRES-Cre mice (available as JAX 024742 and JAX 005628); 3 recordings from a Camk2a-tTA, Ai94 GCaMP6s 2tg × Emx1-IRES-Cre mouse (available as JAX 024115 and JAX 005628); and 2 recordings from a Camk2a-tTA, Ai94 GCaMP6s 2tg × Rasgrf-Cre mouse (available as JAX 024115 and JAX 022864). We also used mice bred to express tdTomato in inhibitory neurons (GAD-IRES-Cre × CAG-tdTomato, available as JAX 010802 and JAX 007909) in 14 recordings. In this case, GCaMP6s was expressed virally, and excitatory neurons were identified by lack of tdTomato expression. These mice were male and female, with ages ranging from 2 months to 8 months. We recorded from sufficient mice to draw scientific conclusions (8 mice in total). There was no randomization or blinding done in the study.

Surgical methods were similar to those described elsewhere[19,36]. In brief, surgeries were performed in adult mice (postnatal day (P)35–P125) under isoflurane anaesthesia (5% for induction, 0.5–1% during the surgery) in a stereotaxic frame. Before surgery, Rimadyl was administered as a systemic analgesic and lidocaine was administered locally at the surgery site. During the surgery we implanted a head-plate for subsequent head-fixation, and made a craniotomy of 3–4 mm in diameter with a cranial window implant for optical access. In Gad-Cre × tdTomato transgenic mice, we targeted virus injections (AAV2/1-hSyn-GCaMP6s, University of Pennsylvania Vector Core, 50–200 nl, 1–3 × 10^{12} GC ml^{-1}) to monocular V1 (2.1–3.3 mm laterally and 3.5–4.0 mm posteriorly from bregma), using a beveled micropipette and a Nanoject II injector (Drummond Scientific Company) attached to a stereotaxic micromanipulator. To obtain large fields of view for imaging, we typically performed 4–8 injections at nearby locations, at multiple depths (around 500 μm and around 200 μm). Rimadyl was then used as a post-operative analgesic for three days, and was delivered to the mice through their drinking water.

Data acquisition. We used a two-photon microscope (Bergamo II multiphoton imaging microscope, Thorlabs) to record neural activity, and ScanImage[37] for data acquisition, obtaining 10,622 ± 1,690 (mean ± s.d.) neurons in the recordings. The recordings were performed using multi-plane acquisition controlled by a resonance scanner, with planes spaced 30–35 μm apart in depth. Ten or twelve planes were acquired sequentially, scanning the entire stack repeatedly at 3 Hz or 2.5 Hz. Because plane scanning was not synchronized to stimulus presentation, we aligned the stimulus onsets to each of the planes separately, and computed stimulus responses from the first two frames acquired after stimulus onset for each plane.

The mice were free to run on an air-floating ball and were surrounded by three computer monitors arranged at 90° angles to the left, front and right of the mouse, so that the head of the mouse was approximately in the geometric centre of the setup. Data from running and non-running periods were analysed together.

For each mouse, recordings were made over multiple days, always returning to the same field of view (in one mouse, two fields of view were used). For each mouse, a field of view was selected on the first recording day such that 10,000 neurons could be observed, with clear calcium transients and a retinotopic location (identified by neuropil fluorescence) localized on the monitors. If more than one potential field of view satisfied these criteria, we chose either a horizontally and vertically central retinotopic location, or a lateral retinotopic location, at 90° from the centre but still centred vertically. The retinotopic location of the field of view (central or lateral) was unrelated to variance spectra. We also did not observe differences between recordings obtained from different modes of GCaMP expression (transgenic versus viral injection). Thus, we pooled data over all conditions.

Visual stimuli. During two-photon recordings, all stimuli other than sparse noise stimuli were presented for 0.5 s, alternating with a grey-screen inter-stimulus interval lasting a random time between 0.3 and 1.1 s. During electrophysiological recordings, all stimuli were presented for 400 ms, with a uniformly distributed inter-stimulus interval of 300–700 ms.

Image stimuli were selected from the ImageNet database[20], from ethologically relevant categories: 'birds', 'cat', 'flowers', 'hamster', 'holes', 'insects', 'mice', 'mushrooms', 'nests', 'pellets', 'snakes' and 'wildcat'. Images were chosen manually to ensure that less than 50% of the image was a uniform background, and to contain a mixture of low and high spatial frequencies. The images were uniformly contrast-normalized. This was achieved by subtracting the local mean brightness and dividing by the local mean contrast (standard deviation of pixel values); the local mean and standard deviation were both computed using a Gaussian filter of standard deviation 30°. Each presented stimulus consisted of a different normalized image from ImageNet (2,800 different images) replicated across all three screens, but at a different rotation on each screen (Fig. 1c).

For the main two-photon recordings, these 2,800 stimuli were presented twice, in the same order each time. In the electrophysiological recordings, 700 of these same stimuli were presented twice in the same order each time. Additionally, in a subset of imaged mice (4 out of 6), we presented a smaller set of 32 images, presented in a randomized order 90–114 times, to enable more accurate estimation of trial-averaged responses.

We also presented partially spatially whitened versions of the 2,800 natural images. To compute spatially whitened images, we first computed the two-dimensional Fourier spectrum for each image, and averaged the spectra across images. We then whitened each image in the frequency domain by dividing its Fourier transform by the averaged Fourier spectrum across all images with a small constant value added for regularization purposes. The rescaled Fourier transform of the image was transformed back into the pixel domain by computing its inverse two-dimensional Fourier transform and retaining the real part. Each image was then intensity-scaled to have the same mean and standard deviation pixel values as the original.

The eight- and four-dimensional stimuli were constructed using a reduced-rank regression model. We first used reduced-rank regression to predict the neuronal population responses R from the natural images I ($N_{pixels} \times N_{stimuli}$) via a d-dimensional bottleneck:

$$R = A^{T} BI$$

where A is a matrix of size $d \times N_{neurons}$ and B is a matrix of size $d \times N_{pixels}$. The dimensionality d was either eight or four depending on the set of stimuli being constructed. The columns of B represent the image dimensions that linearly explain the most variance in the neural population responses. The stimuli were the original 2,800 natural images projected onto the reduced-rank subspace B: $I_{low\text{-}d} = B^{T}BI$.

In addition to natural image stimuli, we also presented drifting gratings and sparse noise. Drifting gratings of 32 directions, spaced evenly at 11.25°, were presented 70–128 times each, lasting 0.5 s each, and with a grey-screen inter-stimulus interval between 0.3 and 1.1 s. They were full-field stimuli (all three monitors) and their spatial frequency was 0.05 cycles per degree and their temporal frequency was 2 Hz.

Sparse noise stimuli consisted of white or black squares on a grey background. Squares were of size 5°, and changed intensity every 200 ms. On each frame, the intensity of each square was chosen independently, as white with 2.5% probability, black with 2.5% probability, and grey with 95% probability. The sparse noise movie contained 6,000 frames, lasting 20 min, and the same movie was played twice to allow cross-validated analysis.

Spontaneous activity was recorded for 30 min with all monitors showing a grey or black background, in all but six of 32 image set recordings. In three recordings of 32-natural image responses and three recordings of drifting grating responses, we interspersed the spontaneous activity, recording 30 s of spontaneous grey-screen activity in between each set of 32 stimuli. In all recordings but these 6, there were also occasional blank stimuli (1 out of every 20 stimuli in the 2,800 natural image stimuli). The activity during these non-stimulus periods was used to project out spontaneous dimensions from the neuronal population responses (see below).

Calcium imaging processing. Calcium movie data was processed using the Suite2p toolbox[19,36], available at https://www.github.com/cortex-lab/Suite2P.

In brief, the Suite2p pipeline consists of registration, cell detection, region of interest (ROI) classification, neuropil correction and spike deconvolution. Movie frames are registered using 2D translation estimated by regularized phase correlation, subpixel interpolation and kriging. To detect ROIs (corresponding to cells), Suite2p clusters correlated pixels, using a low-dimensional decomposition of the data to accelerate processing. The number of ROIs is determined automatically by a threshold on pixel correlations. Finally, ROIs were classified as somatic or non-somatic using a classifier trained on a set of human-curated ROIs. The classifier reached 95% agreement on test data, thus allowing us to skip manual curation for most recordings. For neuropil correction, we used a previously published approach[38], subtracting from each ROI signal the surrounding neuropil signal scaled by a factor of 0.7; all pixels attributed to an ROI (somatic or not) were excluded from the neuropil trace. After neuropil subtraction, we further subtracted a running baseline of the calcium traces with a sliding window of 60 s to remove long-timescale additive baseline shifts in the signals. Fluorescence transients were estimated using non-negative spike deconvolution[39] with a fixed timescale of calcium indicator decay of 2 s, a method that we found to outperform others on ground-truth data[40]. Finally, the deconvolved trace of each cell was z-scored with respect to the mean and standard deviation of the trace of that cell during a 30-min period of grey-screen spontaneous activity before or after the image presentations.

All of the processed deconvolved calcium traces are available on figshare[41] (https://figshare.com/articles/Recordings_of_ten_thousand_neurons_in_visual_cortex_in_response_to_2_800_natural_images/6845348), together with the image stimuli.

Data acquisition and processing (electrophysiology). Neuropixels electrode arrays[42] were used to record extracellularly from neurons in six mice. The mice were between 8 weeks old and 24 weeks old at the time of recording, and were of either sex. The genotypes of the mice were Slc17a7-Cre;Ai95, Snap25-GCaMP6s, TetO-GCaMP6s;CaMKIIa-tTA, Ai32;Pvalb-Cre (two mice), or Emx1-Cre;CaMKIIa-tTA;Ai94. In some cases, other electrophysiological recordings had been made from other locations in the days preceding the recordings reported here. In all cases, a brief (less than 1 h) surgery to implant a steel headplate and 3D-printed plastic recording chamber (12-mm diameter) was first performed. After recovery, mice were acclimatized to head-fixation in the recording setup. During head-fixation, mice were seated on a plastic apparatus with forepaws on a rotating rubber wheel (five mice) or were on a Styrofoam treadmill and able to run (one mouse). Three 20 × 16 cm TFT-LCD screens (LG LP097QX1) were positioned around the mouse at right angles at a distance of 10 cm, covering a total visual angle of 270 × 78 degrees. On the day of recording, mice were again briefly anaesthetized with isoflurane while up to eight small craniotomies were made with a dental drill. After several hours of recovery, mice were head-fixed in the set-up. Probes had a silver wire soldered onto the reference pad and shorted to ground; these reference wires were connected to a Ag/AgCl wire positioned on the skull. The craniotomies as well as the wire were covered with saline-based agar. The agar was covered with silicone oil to prevent drying. Probes were each mounted on a rod held by an electronically positionable micromanipulator (uMP-4, Sensapex) and were then advanced through the agar and through the dura. Once electrodes punctured the dura, they were advanced slowly (10 μm s^{-1}) to their final depth (4 or 5 mm deep). Electrodes were allowed to settle for approximately 15 min before starting recording. Recordings were made in external reference mode with local field potential (LFP) gain = 250 and action potential (AP) gain = 500, using SpikeGLX software. Data were preprocessed by re-referencing to the common median across all channels. Six recordings were performed in six different mice, with a total of 14 probes in visual cortex across all experiments.

We spike-sorted the data using a modification of Kilosort[43] that tracks drifting clusters, called Kilosort2[36,44], available at https://www.github.com/MouseLand/Kilosort2. Without the modifications, the original Kilosort and similar algorithms can split clusters according to drift of the electrode. Kilosort2, in comparison, tracks neurons across drift levels and for longer periods of time (around 1 h in our case).

Removal of ongoing activity dimensions. As shown previously[36], approximately half of the shared variance of visual cortical population activity is unrelated to visual stimuli, but represents behaviour-related fluctuations. This ongoing activity continues uninterrupted during stimulus presentations, and overlaps with stimulus responses only along a single dimension. Because the present study is purely focused on sensory responses, we projected out the dimensions corresponding to ongoing activity before further analysis. The top 32 dimensions of ongoing activity were found by performing a PCA on the z-scored ongoing neural activity recorded during a 30-min period of grey-screen stimuli before or after the image presentations. To remove these dimensions from stimulus responses, the stimulus-driven activity was also first z-scored (using the mean and variance of each neuron computed from spontaneous activity), then the projection onto the 32 top spontaneous dimensions was subtracted (Extended Data Fig. 4).

In the electrophysiological recordings, we considered stimulus responses in a window of 50 ms or 500 ms following stimulus onset. Therefore, we computed the ongoing activity using these two different bin sizes (50 ms or 500 ms). Then we z-scored the stimulus responses by this ongoing activity. Next we computed the top ten PCs of the ongoing activity (in both bin sizes) and then subtracted the projection of the stimulus responses onto these dimensions.

Receptive field estimation. We estimated the receptive fields of the neurons, either using a reduced-rank regression model or using a simple/complex Gabor model. In both cases, the model was fitted to the mean response of each neuron to half of the 2,800 images (I_{train}) over the two repeats. The performance of the model was tested on the mean response of each neuron to the other half of the 2,800 images (I_{test}).

Reduced-rank receptive field estimation. To estimate a linear receptive field for each neuron, we used reduced-rank regression[45], a self-regularizing method that allowed us to fit the responses of all neurons to a single repeat of all 2,800 image stimuli. Reduced-rank regression predicts high-dimensional outputs from high-dimensional inputs through a linear low-dimensional hidden 'bottleneck' representation. We used a 25-dimensional hidden representation to predict the activity of each neuron from the image pixel vectors, taking the resulting regressor matrices as the linear receptive fields. These receptive fields explained 11.4 ± 0.7% (mean ± s.e.m., $n = 7$ recordings) of the stimulus-related variance on the test set. These were z-scored before display in Fig. 1h and Extended Data Fig. 3a.

Model-based receptive field estimation. To fit classical simple/complex receptive fields to each cell, we simulated the responses of a convolutional grid of Gabor filters to the natural images, and fit each neuron with the filter response most correlated to its response.

The Gabor cell filters $G(\boldsymbol{x})$ were parametrized by a spatial frequency f, orientation θ, phase ψ, size α and eccentricity β. Defining $\boldsymbol{u}$ and $\boldsymbol{v}$ to be unit vectors pointing parallel and perpendicular to the orientation θ:

$$G(\boldsymbol{x}) = \cos(2\pi f \boldsymbol{x}\cdot\boldsymbol{u} + \psi)\mathrm{e}^{-((\boldsymbol{x}\cdot\boldsymbol{u})^2+\beta(\boldsymbol{x}\cdot\boldsymbol{v})^2)/2\alpha^2}$$

We constructed 12,288 Gabor filters, with centres spanning a 9 by 7 grid spaced at 5 pixels, and with parameters f, θ, ϕ, α and β ranging from (0.01, 0, 0, 3, 1) to (0.13, 157, 315, 12, 2.5) with (7, 8, 8, 4, 4) points sampled of each parameter, respectively. The parameters were equally spaced along the grid (for example, f was sampled at 0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.13).

Simple cell responses were simulated by passing the dot product of the image with the filter through a rectifier function $r(x) = \max(0, x)$. Complex cell responses were simulated as the root-mean-square response of each unrectified simple cell filter and the same filter with phase ψ shifted by 90°. The activity of a neuron was predicted as a linear combination of a simple cell and its complex cell counterpart, with weights estimated by linear regression. Each neuron was assigned to the filter which best predicted its responses to the training images (Extended Data Fig. 3b–h). This simple/complex Gabor model explained 18.4 ± 0.1% (mean ± s.e.m.) of the stimulus-related variance on the test set.

We also evaluated a model of Gabor receptive fields including divisive normalization[46]. To do so, the response of each of the modelled simple or complex cell filters was divided by the summed, normalized responses of all the other simple and complex cells at this retinotopic location. The experimentally measured response of each neuron was then predicted as a linear combination of simple and complex responses to the best-fitting Gabor, with weights estimated by linear regression. In total, 45.4% ± 1.0% (mean ± s.e.m.) of cells were better fit by the divisive normalization model. However, although divisive normalization changed the optimal parameters fit to many cells (Extended Data Fig. 3i–n), the resulting eigenspectra were indistinguishable from a model with no normalization (Extended Data Fig. 3o–u).

Sparseness estimation. To estimate the sparseness of single-cell responses to the image stimuli, we counted how many neurons were driven more than two standard deviations above their baseline rate by any given stimulus. This was estimated using 4 experiments in which 32 natural images were repeated more than 90 times. We computed the tuning curve of each neuron by averaging over all repeats. The standard deviation of the tuning curve is computed for each neuron across stimuli. The baseline rate was defined as the mean firing rate during all spontaneous activity periods, without visual stimuli. A neuron was judged as responsive to a given stimulus if its response was more than two times this standard deviation plus its baseline firing rate.

Decoding accuracy from 2,800 stimuli. To decode the stimulus identity from the neural responses (Fig. 1g), we built a simple nearest-neighbour decoder based on correlation. The first stimulus presentation was used as the training set while the second presentation was used as the test set. We correlated the population responses for an individual stimulus in the test set with the population responses from all stimuli in the training set. The stimulus with the maximum correlation was then assigned as our prediction. We defined the decoding accuracy as the fraction of correctly labelled stimuli.

Signal-to-noise ratio and explained variance. To compute the tuning-related SNR (Fig. 1f), we first estimated the signal variance of each neuron $\hat{V}_{\text{sig}}$ as the covariance of its response to all stimuli across two repeats (for neuron c, $\hat{V}_{\text{sig}} = \mathrm{Cov}_s[f_1(c,s), f_2(c,s)]$ where $f_r(c,s)$ is the response of neuron c to stimulus s on repeat r, see Supplementary Discussion 1). The noise variance $\hat{V}_{\text{noise}} = V_{\text{tot}} - \hat{V}_{\text{sig}}$ was defined as the difference between the within-repeat variance (reflecting both signal and noise) and this signal variance estimate, and the SNR was defined as their ratio. The SNR estimate is positive when a neuron has responses to stimuli above its noise baseline; note that as $\hat{V}_{\text{sig}}$ is an unbiased estimate, it can take negative values when the true signal variance is zero.

To compute the percentage of explained variance for each neuron (Extended Data Fig. 1c), we divided the estimated signal variance by the total variance across trials (averaged across the repeats):

$$\mathrm{EV} = \frac{\hat{V}_{\text{sig}}}{\frac{1}{2}(\mathrm{Var}_s[f_1(c,s)] + \mathrm{Var}_s[f_2(c,s)])}$$

Note that this formula is similar to the Pearson correlation of the responses of a neuron between two repeats. In the Pearson correlation the numerator is the same, equal to the covariance between repeats, but the denominator is the geometric rather than arithmetic mean of the variances of the two repeats.

cvPCA method. The cvPCA method is fully described in Supplementary Discussion 1, characterized mathematically in Supplementary Discussion 1.1 and 3.6, and analysed in simulation in Extended Data Fig. 5. In brief, the difference between this approach and standard PCA (for example, see previous studies[47,48]) is that it compares the activity on training and test repeats to obtain an estimate

of the stimulus-related ('signal') variance, discounting variance from trial-to-trial variability ('noise').

Denote the response of neuron c to repeat r of stimulus s by $f_r(c, s)$, define the signal as the expected response, which will be equal for both repeats: $\phi(c, s) = \mathbb{E}[f_r(c, s)]$, and the noise on repeat r to be the residual after the expected response is subtracted: $\nu_r(c, s) = f_r(c, s) - \phi(c, s)$. By definition, the noise has zero expectation: $\mathbb{E}_{\nu_r}[\nu_r(c, s)|c, s] = 0$ for all r, c, and s. Let $\hat{\boldsymbol{u}}_n$ denote the nth PC eigenvector, computed from repeat 1.

If we estimated the variance of the projection of activity onto $\hat{\boldsymbol{u}}_n$ using a single repeat, it would contain a contribution from both the signal and the noise. However, because stimulus-independent variability is by definition uncorrelated between repeats of the same stimulus, we can obtain an unbiased estimate of the signal variance, from the covariance across these independent repeats:

$$\mathbb{E}_{\nu_1,\nu_2}\left[\frac{1}{N_s}\sum_{i=1}^{N_s}(\boldsymbol{f}_1(s_i)\cdot\hat{\boldsymbol{u}}_n)(\boldsymbol{f}_2(s_i)\cdot\hat{\boldsymbol{u}}_n)\right]$$

$$=\frac{1}{N_s}\sum_{i=1}^{N_s}(\Phi(s_i)\cdot\hat{\boldsymbol{u}}_n)^2$$

Thus, if $\boldsymbol{u}_n$ is an eigenvector of the population signal variance, the cvPCA method will produce an unbiased estimate of the signal PC variances. As shown in Supplementary Discussion 1.1, this will occur if response variability comprises a mixture of multiplicative response gain changes, correlated additive variability orthogonal to the stimulus dimensions, and uncorrelated noise. Although additive variability in the signal space could in principle downwardly bias the estimated signal variance, other studies confirm that under conditions similar to those analysed here there is little additive variability in the signal space[36]; furthermore, simulations confirm that the amount of such variability present in our recordings does not substantially bias the estimation of signal eigenspectra with cvPCA (Extended Data Fig. 5).

We ran cvPCA ten times on each dataset, on each iteration randomly sampling the population responses of each stimulus from the two repeats without replacement. Thus, $\boldsymbol{f}_1(s)$ could be the population response from either the first or the second repeat, with $\boldsymbol{f}_2(s)$ being the response from the other. The displayed eigenspectra are averages over the ten different runs.

Simulations. To verify that cvPCA method was able to accurately estimate signal eigenspectra in the presence of noise, we analysed simulated data for which the true eigenspectrum was known by construction, and stimulus responses were corrupted by noise. Mathematical analyses (Supplementary Discussion 1.1 and 3.6) showed that noise consisting of multiplicative gain modulation, additive noise orthogonal to signal dimensions, or independent additive noise should not bias the expected eigenspectrum estimate, but that correlated additive noise in the stimulus dimensions could potentially lead the eigenspectrum to be underestimated. We therefore first concentrated on this possibility.

To create the test data, we first simulated noise-free sensory responses, the eigenspectrum of which followed an exact power law, with three possible exponents: $\alpha = 0.5$, 1.0, or 1.5. To simulate the responses of $N_c = 10{,}000$ neurons to $N_s = 2{,}800$ stimuli with this exact eigenspectrum, we first constructed a set of random orthogonal eigenvectors by performing singular value decomposition on a $N_c \times N_s$ matrix A of independent standard Gaussian variates: $A = USV^T$. We created a diagonal matrix D_α, of which the nth diagonal entry was $n^{-\alpha/2}$, and created the $N_c \times N_s$ matrix of simulated noise-free responses as $\phi = UDV^T$.

Additive noise. To simulate correlated additive noise in the stimulus space (Extended Data Fig. 5c), we constructed noise for which the eigenspectrum matched that observed experimentally. To find the empirical noise eigenspectrum, we first estimated the total variance of the nth PC as

$$\hat{\Lambda}_n = \frac{1}{2}\left[\frac{1}{N_s}\sum_{i=1}^{N_s}(\boldsymbol{f}_1(s_i)\cdot\boldsymbol{u}_n)^2 + \frac{1}{N_s}\sum_{i=1}^{N_s}(\boldsymbol{f}_2(s_i)\cdot\boldsymbol{u}_n)^2\right]$$

and estimated the signal variance using cvPCA as

$$\hat{\lambda}_n = \frac{1}{N_s}\sum_{i=1}^{N_s}(\boldsymbol{f}_1(s_i)\cdot\boldsymbol{u}_n)(\boldsymbol{f}_2(s_i)\cdot\boldsymbol{u}_n)$$

The estimated noise spectrum was the difference between total variance and signal variance: $\hat{\delta}_n = \hat{\Lambda}_n - \hat{\lambda}_n$. This spectrum reflects the summed magnitude of both correlated and uncorrrelated noise in the signal dimensions, and is shown in Extended Data Fig. 5b. Responses corrupted by additive noise were simulated as $\Phi + b_\alpha U\Delta V^T$, where Δ is a diagonal matrix with entries δ_n, and the scale factor b_α ensured that, as in the data, the simulation showed a total of 14% reliable variance. The scale factors were found by search to be 2.62, 2.52, 2.41 for the signal eigenspectrum exponents $\alpha = 0.5$, 1.0, 1.5, respectively.

Multiplicative noise. To simulate multiplicative noise (Extended Data Fig. 5d), responses were multiplied by an amplitude factor that was constant across neurons, but was drawn independently for each stimulus and repeat. To simulate an appropriately skewed distribution of gains, the scale factor was distributed as 0.5 plus an exponential random variate with mean parameter c_α. The values of c_α were found by search as those matching the observed 14% reliable variance, yielding 1.55, 1.52, 1.40 for the signal eigenspectrum exponents $\alpha = 0.5$, 1.0, 1.5, respectively.

To simulate a combination of additive and multiplicative noise (Extended Data Fig. 5e), responses were modulated by the additive mechanism described above and then modulated multiplicatively. The gain factors were $b_\alpha = 0.55$, 0.53, 0.51 and $c_\alpha = 0.65$, 0.64, 0.59 for $\alpha = 0.5$, 1.0, 1.5 respectively.

Two-photon noise. To investigate whether our two-photon deconvolution method could be biasing the estimated eigenspectrum, we simulated the effect of passing noise through this algorithm (Extended Data Fig. 5f).

To do so, we extended the simulations above to apply in the time domain. When simulating the additive noise, we allowed it to vary across all simulated two-photon imaging frames (replacing the matrix A used to compute the eigenvectors U and V by a 10,000 × 8,400 matrix providing three simulated frames per stimulus presentation). The gain modulation factor was assumed equal for all three frames corresponding to a single stimulus. The magnitudes of the additive noise and the gain factor giving 14% signal variance were found by search to be $b_\alpha = 0.50$, 0.50, 0.49 and $c_\alpha = 0.68$, 0.67, 0.66, for $\alpha = 0.5$, 1.0, 1.5, respectively.

To simulate the response of GCaMP6s, we convolved these responses with an exponentially decaying kernel with a timescale of 2 s (because each time point in the data is 0.4 s, this corresponds to a decay timescale of five time points). To simulate shot noise, we added Gaussian white noise with a standard deviation of 0.5. Next we deconvolved these noisy traces using OASIS[39], with a timescale of 5 time points and no sparsity constraints. The reduction in signal variance from this procedure was roughly 1%.

For all noise simulations, we estimated the signal eigenspectrum from two repeats using cvPCA. We found that cvPCA, but not ordinary PCA, correctly estimated the ground-true eigenspectrum, for all simulated power-law exponents α (Extended Data Fig. 5g).

Estimation of power-law exponent. We computed the linear fit of the eigenspectrum over the range of 11 to 500 dimensions for all recordings (and model fits) other than the 32 drifting grating recordings. For the 32-grating recordings, owing to noise and the length of the spectrum, we computed the power-law exponent from 5 to 30. The linear fit was performed in log–log space: the range of log(11) to log(500) was regressed onto the log of the eigenspectrum, sampled at points that were themselves logarithmically spaced between 11 and 500.

Sorting neurons and stimuli by correlations. In Extended Data Fig. 6, neurons and stimuli were sorted so that they were close to other neurons and stimuli with which they were correlated.

To do this, we first z-scored the binned activity of each neuron and computed PCs of its averaged activity across repeats. Each panel shows this for different PC projections of the data: 1, 2, 3–10, 11–40, 41–200 and 201–1,000. Stimuli were re-ordered so that the pattern of evoked population activity of each stimulus was most similar to the average of its neighbours. The stimulus order was initialized by sorting stimuli according to their weights on the top PC of activity, then dividing them into 30 clusters of equal size along this ordering. For 50 iterations, we computed the mean activity of each cluster and smoothed this activity across clusters with a Gaussian, the width of which was annealed from 6 clusters to 1 over the first 25 iterations. Each stimulus was then reassigned to the cluster it was most correlated with. On the final pass, we upsampled the correlations of the stimuli with each cluster by a factor of 100 using kriging interpolation (smoothing constant of 1 cluster), resulting in a continuous assignment of stimuli along the 1D axis of the clustering algorithm. After sorting across stimuli, we smoothed across them to reduce noise, recomputed the PCs on the activity smoothed across stimuli, and repeated the procedure to sort neurons. The algorithm is available in Python and MATLAB at https://www.github.com/MouseLand/RasterMap. These plots were made using the MATLAB version of the code.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

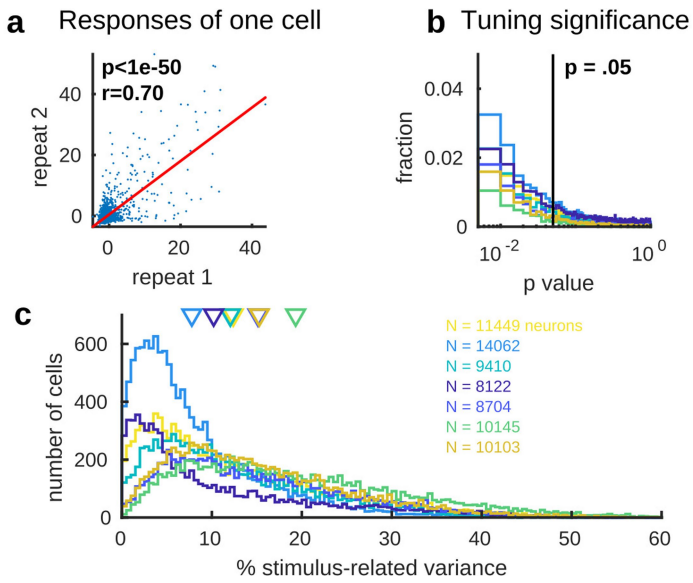
Data availability

All of the processed deconvolved calcium traces are available on figshare[41] (https://figshare.com/articles/Recordings_of_ten_thousand_neurons_in_visual_cortex_in_response_to_2_800_natural_images/6845348), together with the image stimuli.

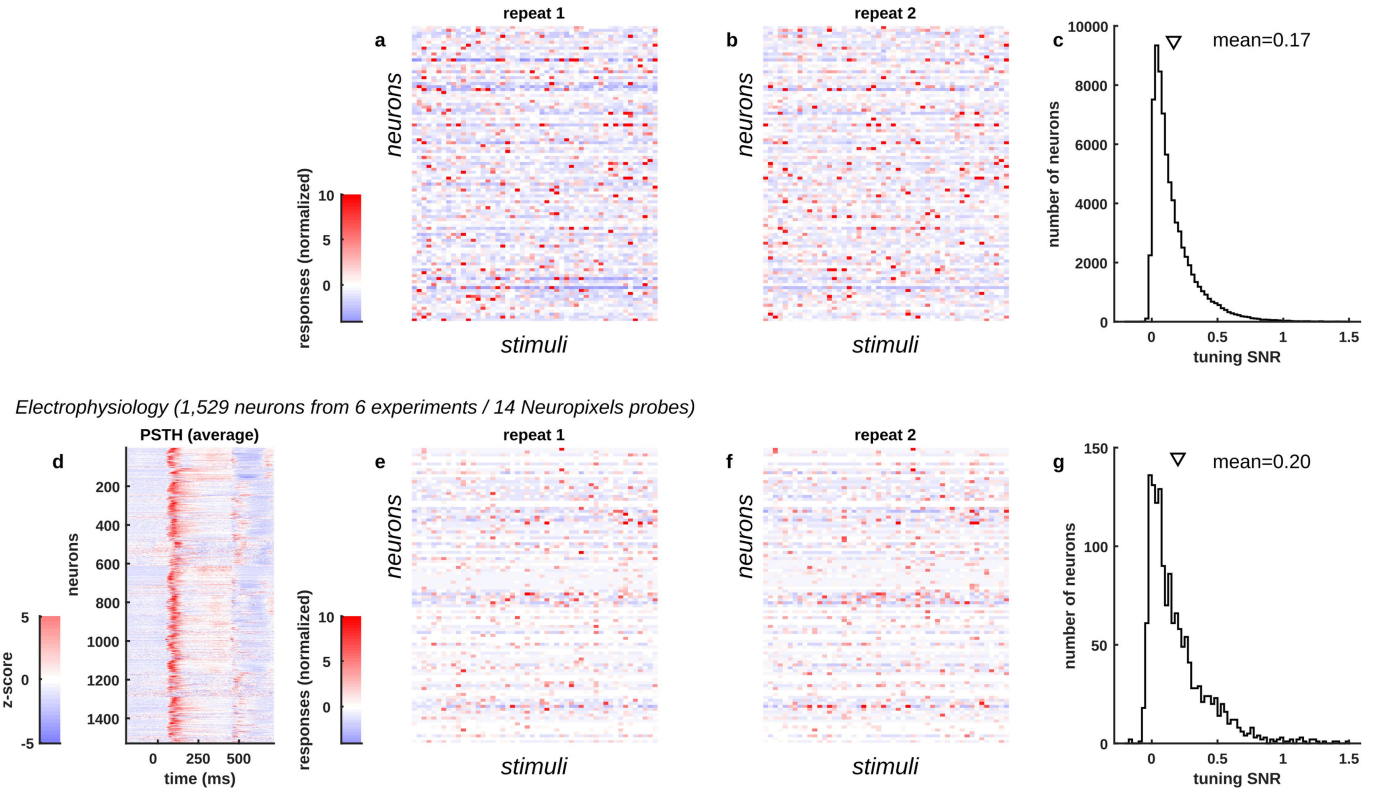
Code availability

The code is available on GitHub (https://github.com/MouseLand/stringer-pachitariu-et-al-2018b).

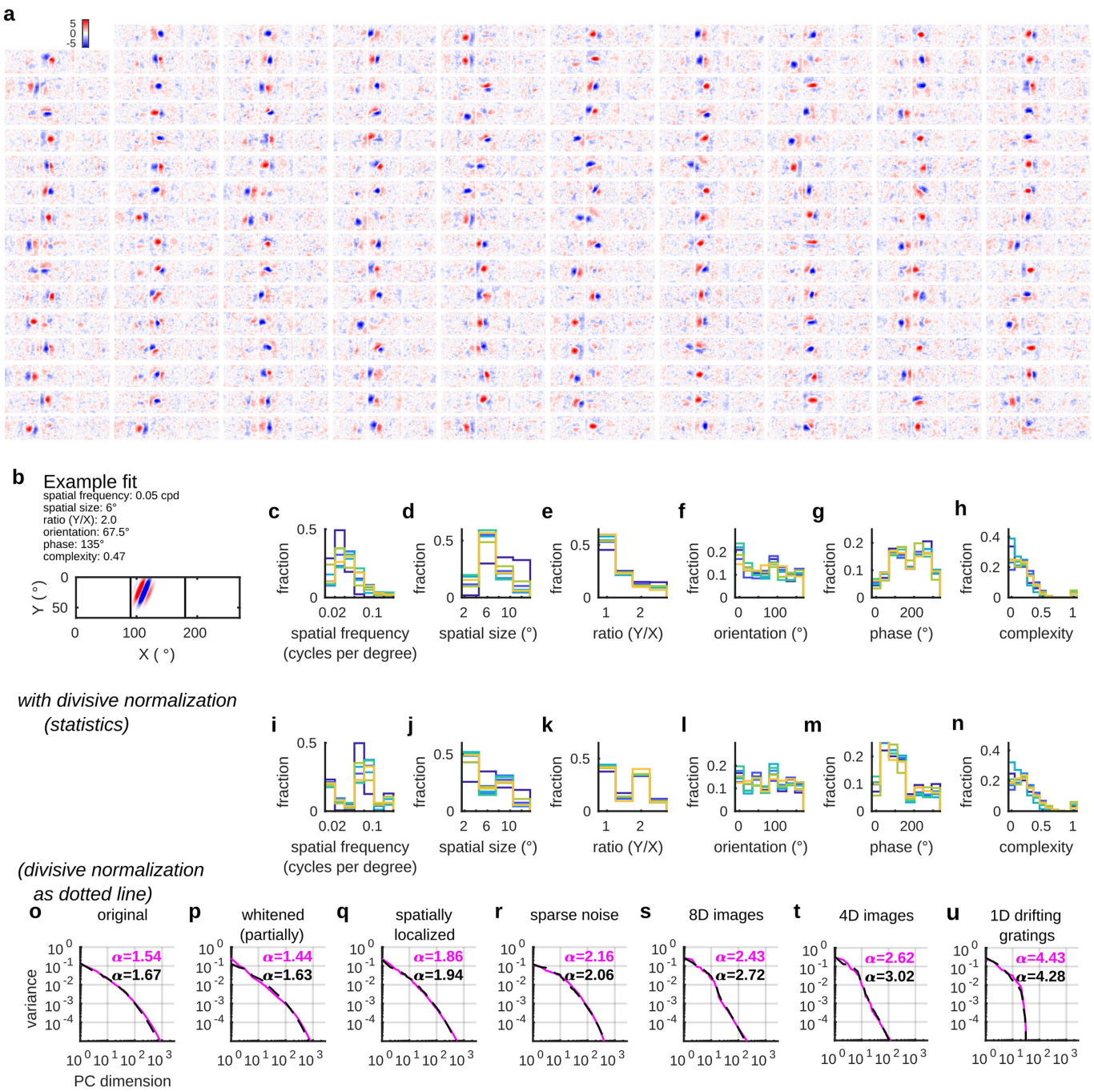
36. Stringer, C. et al. Spontaneous behaviors drive multidimensional, brainwide activity. *Science* **364**, eaav7893 (2019).
37. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
38. Chen, T.-W. et al. Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295–300 (2013).
39. Friedrich, J., Zhou, P. & Paninski, L. Fast online deconvolution of calcium imaging data. *PLoS Comput. Biol*. **13**, e1005423 (2017).
40. Pachitariu, M., Stringer, C. & Harris, K. D. Robustness of spike deconvolution for neuronal calcium imaging. *J. Neurosci*. **38**, 7976–7985 (2018).
41. Stringer, C., Pachitariu, M., Carandini, M. & Harris, K. Responses of ten thousand neurons to 2,800 natural images. *Figshare* https://doi.org/10.25378/janelia.6845348.v3 (2018).
42. Jun, J. J. et al. Fully integrated silicon probes for high-density recording of neural activity. *Nature* **551**, 232–236 (2017).
43. Pachitariu, M., Steinmetz, N. A., Kadir, S. N., Carandini, M. & Harris, K. D. Fast and accurate spike sorting of high-channel count probes with kilosort. In *Advances in Neural Information Processing Systems* 4448–4456 (2016).
44. Allen, W. E. et al. Thirst regulates motivated behavior through modulation of brainwide neural population dynamics. *Science* **364**, eaav3932(2019).
45. Izenman, A. J. Reduced-rank regression for the multivariate linear model. *J. Multivariate Anal*. **5**, 248–264 (1975).
46. Schwartz, O. & Simoncelli, E. P. Natural signal statistics and sensory gain control. *Nat. Neurosci*. **4**, 819–825 (2001).
47. Cowley, B. R., Smith, M. A., Kohn, A. & Yu, B. M. Stimulus-driven population activity patterns in macaque primary visual cortex. *PLoS Comput. Biol*. **12**, e1005185 (2016).
48. Gallego, J. A. et al. Cortical population activity within a preserved neural manifold underlies multiple motor behaviors. *Nat. Commun*. **9**, 4233 (2018).



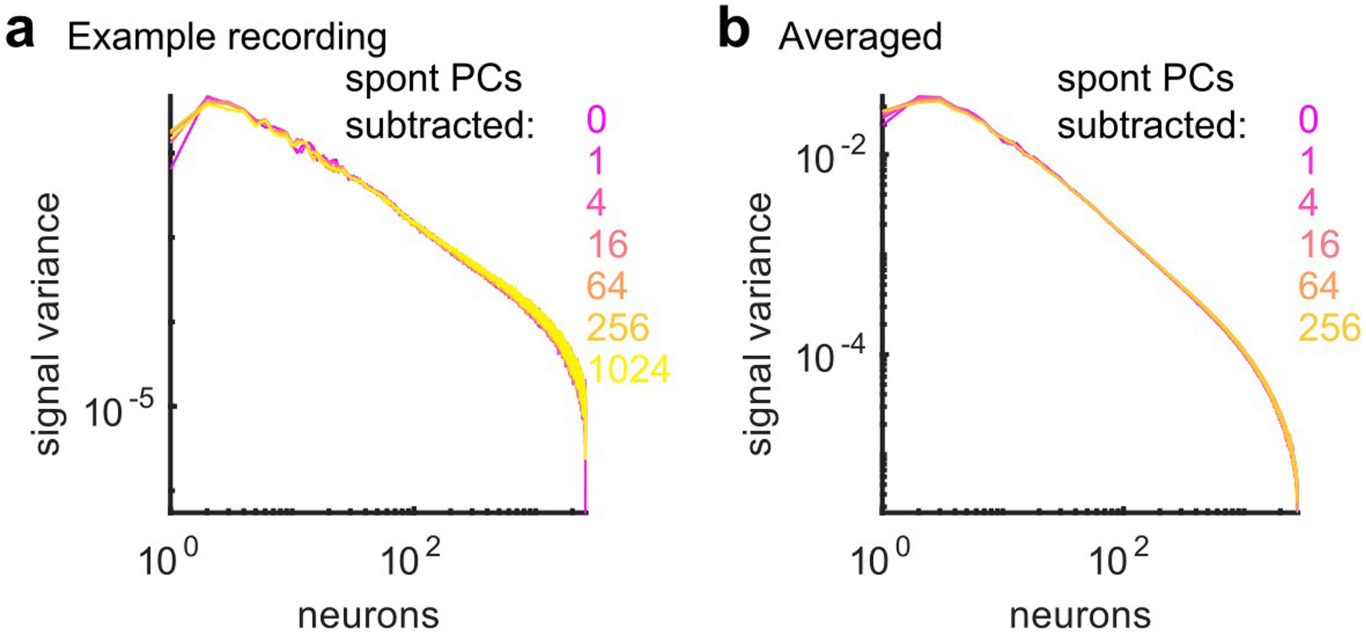
Extended Data Fig. 1 | Reliability of single-neuron responses. **a**, The responses of a single neuron to the first repeat of 2,800 stimuli plotted against its responses to the second repeat of the same stimuli. **b**, Histograms of *P* values for Pearson correlation of responses on the two repeats. Each coloured histogram represents a different recording. In total, 81.4 ± 5.1% (mean ± s.e.m., $n = 7$ recordings) of cells were significant at $P < 0.05$. **c**, Histogram of the single-neuron percentage of stimulus-related variance across the population. Each coloured histogram represents a different recording; arrowheads (top) represent the mean for each experiment.



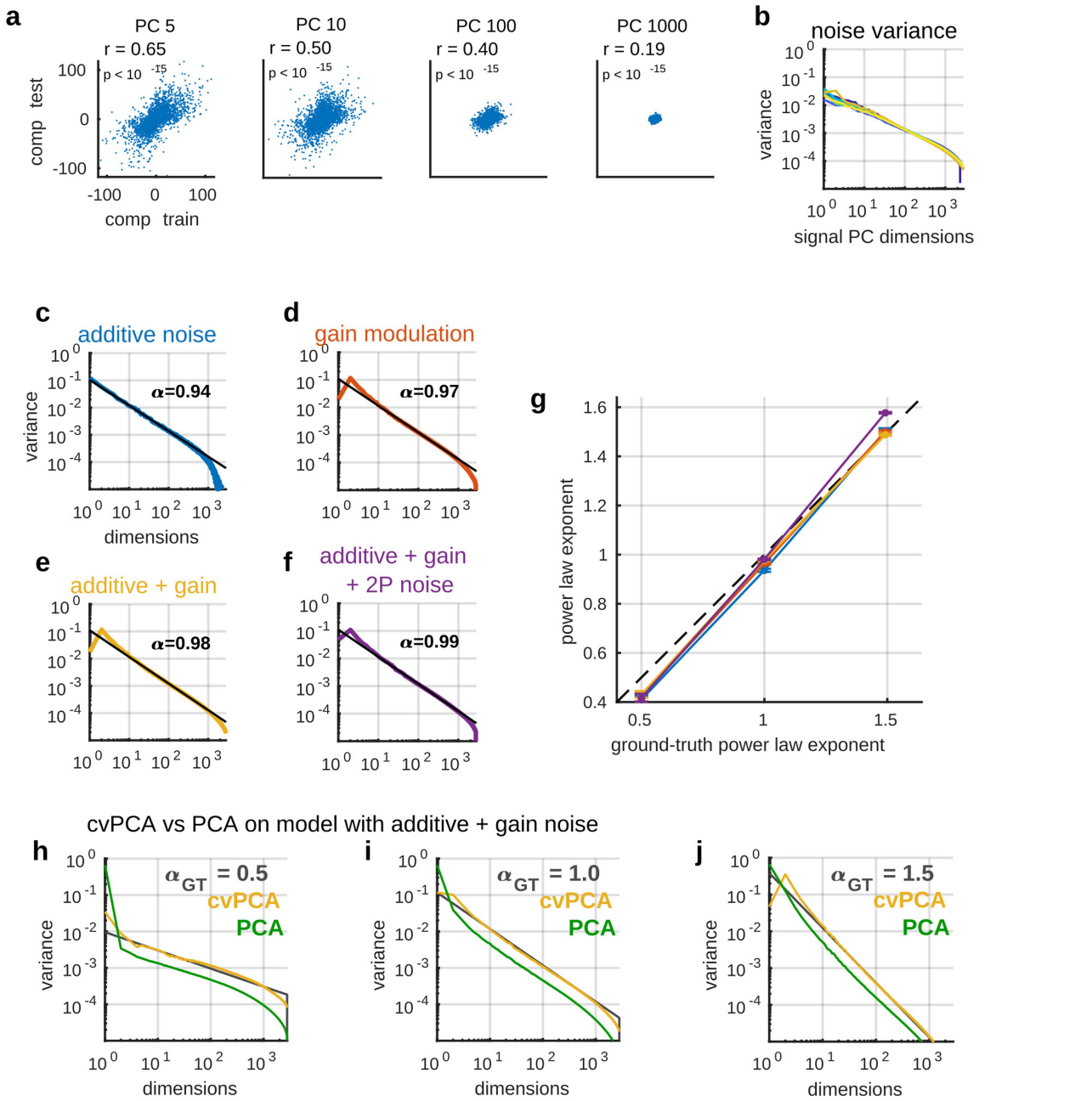
Extended Data Fig. 2 | Comparison with electrophysiology. a, **b**, Single trial responses of 100 neurons to two repeats of 50 stimuli, recorded by two-photon calcium imaging. **c**, Distribution of tuning SNR for 74,353 neurons recorded by two-photon calcium imaging. **d**, Average peristimulus time histogram of spikes recorded electrophysiologically in a separate set of experiments. The images shown were a random subset of 700 images out of the total 2,800. The peristimulus time histogram reflects the average over all stimuli. The responses are *z*-scored across time for each neuron. **e**–**g**, Same as **a**–**c** for the electrophysiologically recorded neurons.



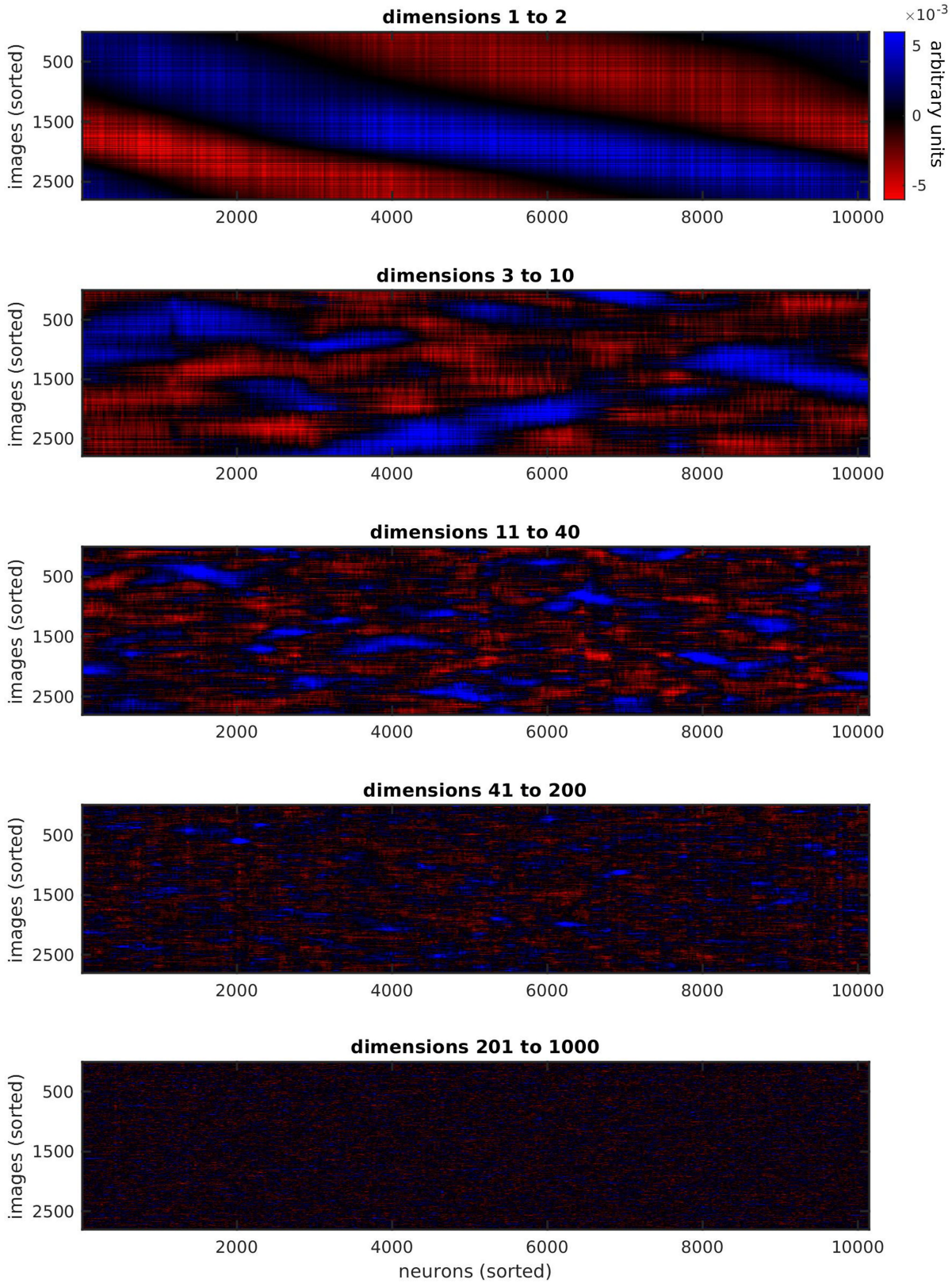
Extended Data Fig. 3 | Single-neuron receptive fields estimated using reduced-rank regression and Gabor models. a, The receptive fields of 159 randomly chosen neurons, estimated using reduced-rank regression. The receptive field map is *z*-scored for each neuron. **b**, An example Gabor fit to a single cell. **c**–**h**, Histograms showing the distribution of model parameters across cells. Each colour represents cells from one recording. **i**–**n**, Histograms showing the distribution of model parameters across cells when the model also has divisive normalization. **o**–**u**, Eigenspectra of Gabor population model responses to the different stimulus sets, as labelled. The unnormalized Gabors are shown in magenta, and the model with divisive normalization in black.



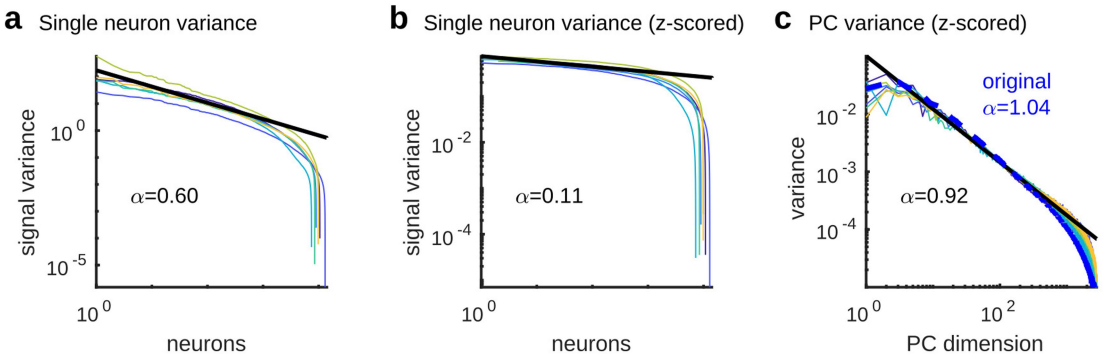
Extended Data Fig. 4 | Stimulus-independent activity does not affect the measured eigenspectrum. a, To measure the effects of correlated noise variability on eigenspectra estimated by cvPCA, we examined the effect of projecting out different numbers of noise dimensions (estimated during periods of spontaneous grey-screen) from the responses in an example experiment. **b**, The same analysis as in **a**, averaged over all recordings. The presence of these noise dimensions made little difference to the estimated signal eigenspectrum other than to slightly reduce estimated eigenvalues in the highest and lowest dimensions. For the main analyses, 32 spontaneous dimensions were subtracted.



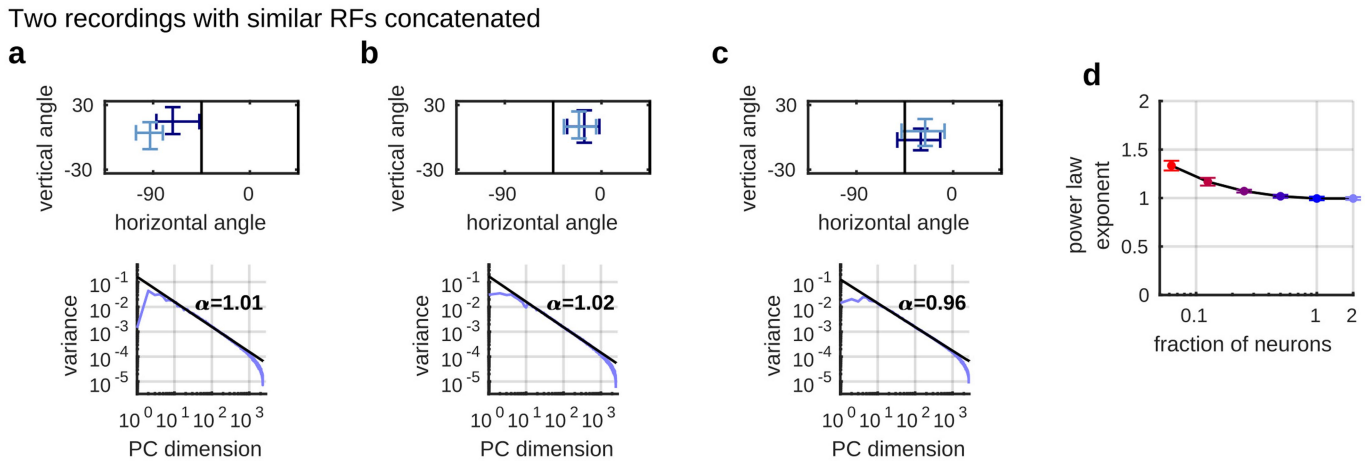
Extended Data Fig. 5 | Validating the eigenspectrum estimation method using simulations with the true noise distribution. a, Scatterplots illustrating the noise levels of each estimated PC. Each plot shows population activity projected onto the specified PC, for the first repeat (x axis) and second repeat (y axis). Each point represents responses to a single stimulus. **b**, Estimated level of noise variance in successive signal dimensions. Noise variance was estimated by subtracting the cvPCA estimate of signal variance from the total variance (see Methods). **c**, Recovery of ground-truth eigenspectrum in simulated data. We simulated responses of 10,000 neurons to 2,800 stimuli with a power spectrum decay of exactly $\alpha = 1$, and added noise in the stimulus space, generated with the spectrum in **b** scaled to produce the same signal-to-noise ratio as in the original neural data. The ground-truth eigenspectrum (black) is estimated accurately by the cvPCA method (blue). **d**, Same analysis as in **c** with multiplicative noise, in which the responses of all neurons on each trial were multiplied by a common random factor. The distribution of this factor was again scaled to recover the original signal-to-noise ratio. **e**, Same analysis as in **c** with a combination of multiplicative and additive noise. **f**, Same analysis as in **c**, also including simulation of neural and two-photon shot noise before running a GCaMP deconvolution algorithm. **g**, Ten instantiations of the simulation were performed with ground-truth exponents of 0.5, 1.0 and 1.5. Error bars represent standard deviations of the power-law exponents estimated for each of the ten simulations. The dashed black line represents the ground-truth value. **h**–**j**, Comparison of cvPCA (yellow) and traditional PCA (green) algorithms in the presence of the additive + multiplicative noise combination. Whereas cvPCA recovered the ground-truth eigenspectrum (black) exactly, traditional PCA did not, resulting in overestimation of the top eigenvalues and failure to detect the ground-truth power law.



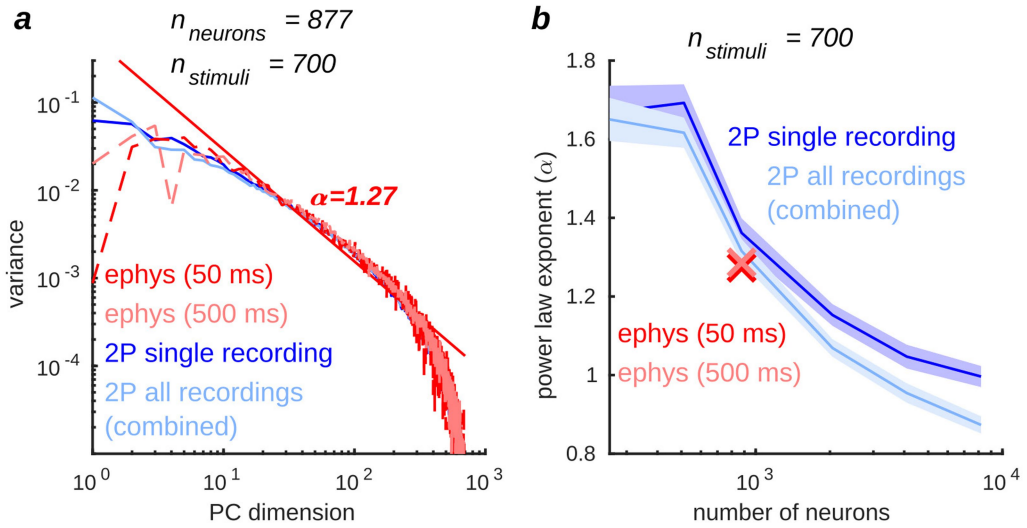
Extended Data Fig. 6 | Successive PC dimensions encode finer stimulus features. Each plot shows the responses of 10,145 neurons to 2,800 natural images, projected onto the specified PCs and then sorted along both axes so that correlated neurons and stimuli are close together. We then smoothed the matrix across neurons and stimuli with Gaussian kernels of widths of 8 neurons and 2 stimuli, respectively. Dimensions 1–2 reveal a coarse, one-dimensional organization of the neurons and stimuli. Dimensions 3–10 reveal multidimensional structure, which involves different neural subpopulations responding to different stimuli. Dimensions 11–40 reveal finer-structured patterns of correlated selectivity among neurons. Dimensions 41–200 and 201–1,000 reveal even finer-structured selectivity, which contained less neural variance.



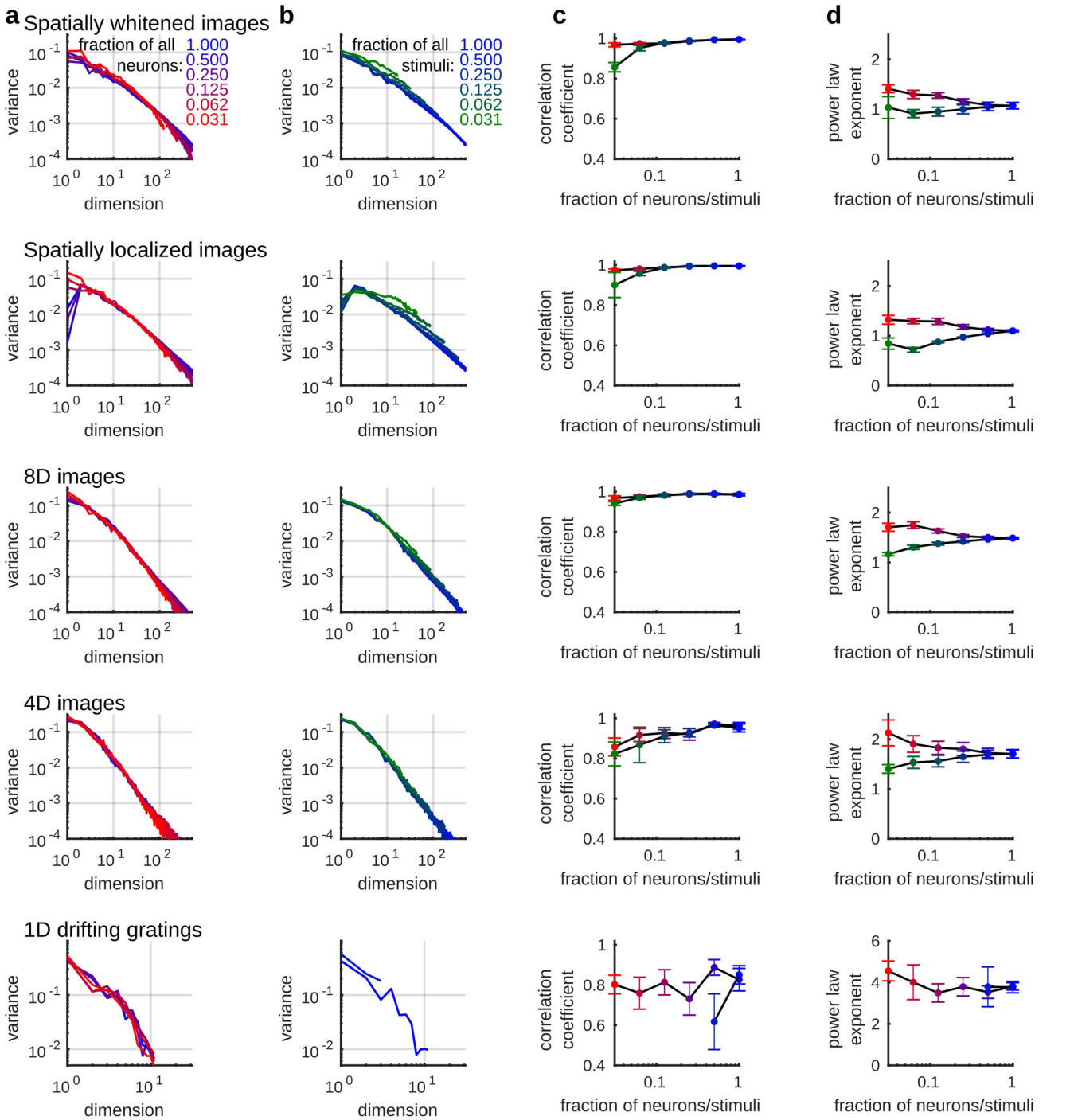
Extended Data Fig. 7 | Power-law scaling reflects correlation structure, not single-neuron statistics. a, The signal variance of the responses of each neuron are sorted in descending order; they approximately follow a power law with a decay exponent of $\alpha = 0.59$. **b**, The same plot after *z*-scoring the recorded traces to equalize stimulus response sizes between cells; the distribution of single-neuron variance has become nearly flat. **c**, PC eigenspectra for *z*-scored data. Each coloured line represents a different recording. The dashed blue line shows the average eigenspectrum from the original, non-*z*-scored responses. The fact that the eigenspectrum power law is barely affected by equalizing firing rates, whereas the distribution of single-cell signal variance is altered, indicates that the power law arises from correlations between cells rather than from the distribution of firing rates or signal variance across cells.



Extended Data Fig. 8 | Power-law eigenspectra in concatenated recordings. a–c, To investigate whether power-law eigenspectra apply to even larger populations, we were able to artificially double the number of recorded neurons by combining three pairs of recordings for which the imaging fields of view had similar retinotopic locations. Top, retinotopic locations of receptive fields (95% confidence intervals on the mean receptive field position of that recording), with each recording shown in a different shade of blue. Bottom, eigenspectrum of concatenated recordings in response to the 2,800 natural image stimuli; total population sizes 19,571, 23,472 and 18,807 cells respectively. Each panel (**a**, **b** and **c**) represents one pair of recordings. **d**, Eigenspectrum exponents for random subsets of the combined populations (compare with Fig. 2j). The horizontal axis shows the population size relative to single recordings, so the merged population has size 2. The mean power-law exponent for fraction of neurons = 2 was $\alpha = 0.99 \pm 0.02$ (mean ± s.e.m.).



Extended Data Fig. 9 | Eigenspectrum of electrophysiologically recorded data. We recorded neural activity electrophysiologically in response to 700 out of the 2,800 stimuli, and concatenated the recordings, resulting in a total of 877 neurons recorded across 6 experiments. **a**, With this smaller number of stimuli and neurons, convergence to a power law is not complete, and the exponent cannot be estimated accurately (compare with Fig. 2g–j). We therefore compared the electrophysiology data to the responses generated by these stimuli in 877 neurons sampled randomly from either a single two-photon imaging experiment (dark blue) or all experiments combined (light blue). The red and pink colours show electrophysiology eigenspectra with time bins of 50 ms or 500 ms; the red line shows the best linear fit to estimate the exponent. **b**, The blue curves represent power-law exponents estimated from the responses of different-sized neuronal subpopulations to this set of 700 stimuli; the shading represents s.e.m. over different random subsets of neurons. The red and pink crosses denote estimated exponents from electrophysiology data for 50-ms and 500-ms bin sizes.



Extended Data Fig. 10 | Power-law scaling grows more accurate for increasing numbers of neurons and stimuli, for all stimulus ensembles. **a**, Eigenspectra estimated from a random subset of the recorded neurons, colour-coded by the fraction of neurons retained. **b**, Eigenspectra estimated from a random subset of stimuli, colour-coded by the fraction of stimuli retained. **c**, Correlation coefficient of the spectra plotted in **a**, **b**. **d**, Power-law exponent of the spectra plotted in **a**, **b**. Each row corresponds to a different ensemble of visual stimuli.

nature research

Corresponding author(s): Carsen Stringer, Marius Pachitariu, Kenneth Harris

Last updated by author(s): Apr 22, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	We used the open-source Scanimage v4.2 software for two-photon calcium imaging data acquisition (http://scanimage.vidriotechnologies.com/display/SIH/ScanImage+Home). Electrophysiological acquisition was done using the open-source SpikeGLX software (https://billkarsh.github.io/SpikeGLX/).
Data analysis	We processed the two-photon calcium imaging data using suite2p (github.com/MouseLand/suite2p) and the electrophysiological data using Kilosort2 (github.com/MouseLand/kilosort2). The one-dimensional visualization used rastermap (github.com/MouseLand/rastermap). All the code to reproduce the figures is provided at github.com/MouseLand/stringer-pachitariu-et-al-2018b.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All the processed two-photon calcium imaging is provided on figshare: https://figshare.com/articles/Recordings_of_ten_thousand_neurons_in_visual_cortex_in_response_to_2_800_natural_images/6845348.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The sample size was sufficiently large (8 mice from which 42 recordings containing ~10,000 neurons each were performed).
Data exclusions	We excluded two recordings which did not have significant signal variance, suggesting we were not recording in primary visual cortex.
Replication	We observed similar results across mice in terms of the power law decay of signal variance across dimensions.
Randomization	This is not relevant to our study because it is an observational study looking at the activity of neurons in visual cortex in healthy awake adult mice.
Blinding	Blinding was not relevant to our study (see above).

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	We used mice bred to express GCaMP6s in excitatory neurons in our recordings: 13 recordings from TetO-GCaMP6s x Emx1-IRES-Cre mice (available as JAX 024742 and JAX 005628); 3 recordings from a Camk2a-tTA, Ai94 GCaMP6s 2tg x Emx1-IRES-Cre mouse (available as JAX 024115 and JAX 005628); and 2 recordings from a Camk2a-tTA, Ai94 GCaMP6s 2tg x Rasgrf-Cre mouse (available as JAX 024115 and JAX 022864). We also used mice bred to express tdTomato in inhibitory neurons (GAD-IRES-Cre x CAG-tdTomato, available as JAX 010802 and JAX 007909) in 14 recordings. These mice were male and female, and ranged from age 2 to 8 months.
Wild animals	*Provide details on animals observed in or captured in the field; report species, sex and age where possible. Describe how animals were caught and transported and what happened to captive animals after the study (if killed, explain why and describe method; if released, say where and when) OR state that the study did not involve wild animals.*
Field-collected samples	*For laboratory work with field-collected samples, describe all relevant parameters such as housing, maintenance, temperature, photoperiod and end-of-experiment protocol OR state that the study did not involve samples collected from the field.*
Ethics oversight	All experimental procedures were conducted according to the UK Animals Scientific Procedures Act (1986). Experiments were performed at University College London under personal and project licenses released by the Home Office following appropriate ethics review.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

ARTICLE

https://doi.org/10.1038/s41586-019-1344-7

Structure and autoregulation of a P4-ATPase lipid flippase

Milena Timcenko[1,5], Joseph A. Lyons[1,5], Dovile Januliene[2,5], Jakob J. Ulstrup[1], Thibaud Dieudonné[3], Cédric Montigny[3], Miriam-Rose Ash[1], Jesper Lykkegaard Karlsen[1], Thomas Boesen[1,4], Werner Kühlbrandt[2], Guillaume Lenoir[3]*, Arne Moeller[2]* & Poul Nissen[1]*

Type 4 P-type ATPases (P4-ATPases) are lipid flippases that drive the active transport of phospholipids from exoplasmic or luminal leaflets to cytosolic leaflets of eukaryotic membranes. The molecular architecture of P4-ATPases and the mechanism through which they recognize and transport lipids have remained unknown. Here we describe the cryo-electron microscopy structure of the P4-ATPase Drs2p–Cdc50p, a *Saccharomyces cerevisiae* lipid flippase that is specific to phosphatidylserine and phosphatidylethanolamine. Drs2p–Cdc50p is autoinhibited by the C-terminal tail of Drs2p, and activated by the lipid phosphatidylinositol-4-phosphate (PtdIns4P or PI4P). We present three structures that represent the complex in an autoinhibited, an intermediate and a fully activated state. The analysis highlights specific features of P4-ATPases and reveals sites of autoinhibition and PI4P-dependent activation. We also observe a putative lipid translocation pathway in this flippase that involves a conserved PISL motif in transmembrane segment 4 and polar residues of transmembrane segments 2 and 5, in particular Lys1018, in the centre of the lipid bilayer.

Cells and organelles are defined by lipid-bilayer membranes and by membrane proteins. In eukaryotic membranes that are involved in the late secretory and endocytic pathways, the lipid distributions between the two bilayer leaflets are asymmetric. These lipid gradients potentiate key biological processes such as membrane dynamics, endocytosis and exocytosis, and signalling[1–4]. Owing to membrane-fusion events and the bidirectional and gradient-dissipating activity of lipid scramblases, lipid asymmetry must constantly be regulated and restored. Members of two distinct superfamilies of membrane proteins drive the ATP-dependent unidirectional translocation of lipids against concentration gradients. ATP-binding cassette (ABC) transporters typically drive the inward-to-outward (flop) translocation of lipids between bilayer leaflets, whereas P4-ATPases drive the outward-to-inward (flip) process[1–4]. The P-type ATPases couple transport to the formation and breakdown of the phosphoenzyme through a functional cycle that involves several intermediate states (E1, E1P, E2P and E2) (Extended Data Fig. 1a). P4-ATPases couple lipid transport to dephosphorylation of the E2P state[5,6], in a manner that is similar to the inward transport of potassium by the Na^+, K^+-ATPase. Conversely, phosphorylation of E1P appears to be independent of transport[5,7] (Extended Data Fig. 1a, b). Although previous work has shed light on the structure and mechanism of lipid floppases[8,9] and scramblases[10,11], P4-ATPases have so far been studied only through bioinformatics and functional assays. Models have identified potential peripheral or centrally located lipid-recognition sites and pathways[12–14], but the mechanism of transport remains a source of debate.

Most P4-ATPases are binary complexes that contain a subunit from the CDC50 family of proteins; this subunit is necessary for the correct localization and function of the complex[15,16] (Extended Data Fig. 1c). Mutant forms of mammalian lipid flippases have been implicated in various diseases; for example, ATP8A1 and ATP8A2 in neurological disorders, ATP8B1 in progressive familial intrahepatic cholestasis type 1, ATP10A in type 2 diabetes and insulin resistance and ATP11A in cancer[17]. The *trans*-Golgi-localized Drs2p–Cdc50p complex from the yeast *S. cerevisiae* is well-characterized. In vivo[18,19] and in vitro[20,21] studies have shown that Drs2p–Cdc50p primarily flips phosphatidylserine (and—to a lesser extent—phosphatidylethanolamine) from the luminal side to the cytosolic leaflet, and indicate that this function may have a role in the biogenesis of vesicles at late secretory membranes[22].

The C terminus of Drs2p contains an autoinhibitory domain[23,24], and relief of autoinhibition requires the regulatory lipid PI4P[6,23]. Binding of Gea2p (a guanine nucleotide exchange factor for the small GTPase Arf) to a basic segment of the C terminus has previously been reported to be necessary for activation of Drs2p in vivo[23], although this finding has not yet been confirmed in vitro[24]. Furthermore, interaction of Arl1p (a GTPase of the Arf family) with the extended N terminus of Drs2p has been implicated in flippase activity in vivo[25]. The first 104 amino acids of the N terminus have little effect on in vitro activity; by contrast, truncation of the C terminus is activating, but the protein continues to be under regulation by PI4P[26] (Extended Data Fig. 1d). Although these studies highlight the components that are involved, a molecular mechanism of autoregulation for Drs2p–Cdc50p (and P-type ATPases in general) remains unknown.

To gain structural and mechanistic insights, we embarked on cryo-electron microscopy (cryo-EM) studies of Drs2p–Cdc50p complexes, stabilized with beryllium fluoride (BeF_3^-). Drs2p–Cdc50p was overexpressed in *S. cerevisiae* and purified in lauryl maltose neopentyl glycol (LMNG) detergent by affinity chromatography and gel filtration, resulting in a monodisperse sample that contained both subunits of the complex (Extended Data Fig. 2). The resulting structures are in conformations that are consistent with an E2P phosphoenzyme and capture the progressive steps from a fully autoinhibited P4-ATPase ($E2P^{inhib}$), to an intermediate activated state in the presence of PI4P ($E2P^{inter}$), and an outward-open and activated conformation that is represented by a C-terminally truncated enzyme, also in the presence of PI4P ($E2P^{active}$).

[1]DANDRITE, Nordic EMBL Partnership for Molecular Medicine, Department of Molecular Biology and Genetics, Aarhus University, Aarhus, Denmark. [2]Max Planck Institute for Biophysics, Frankfurt, Germany. [3]Institute for Integrative Biology of the Cell (I2BC), CEA, CNRS, Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France. [4]Interdisciplinary Nanoscience Center (iNANO), Aarhus University, Aarhus, Denmark. [5]These authors contributed equally: Milena Timcenko, Joseph A. Lyons, Dovile Januliene. *e-mail: guillaume.lenoir@i2bc.paris-saclay.fr; arne.moeller@biophys.mpg.de; pn@mbg.au.dk

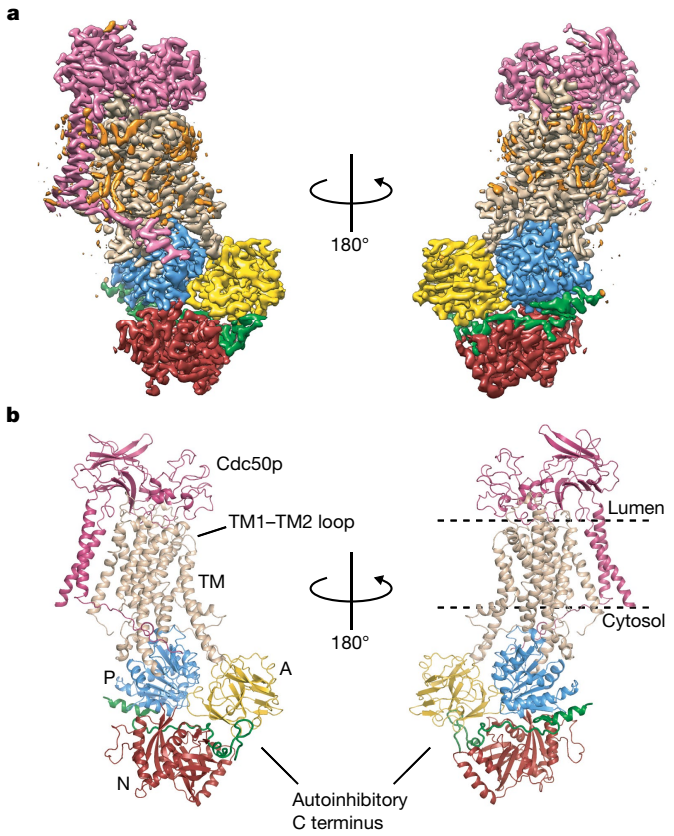


Fig. 1 | Architecture of the Drs2p–Cdc50p complex. a, LocScale map[35] of E2P^{inhib}, coloured by domain: the A, P and N domains of Drs2p are yellow, blue and red, respectively; the transmembrane domain is tan; the autoinhibitory C terminus is green; and Cdc50p is pink. Unmodelled map features that correspond to ordered lipid or detergent molecules are orange. **b**, Cartoon representation of the refined E2P^{inhib} model. Colours are as in **a**.

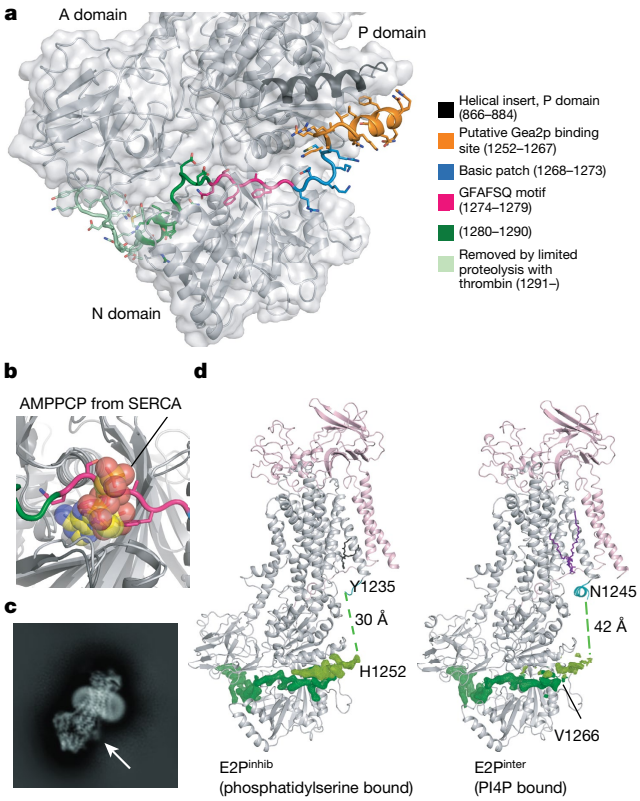


Fig. 2 | Autoinhibition of Drs2p by its C terminus. a, The autoinhibitory domain bound between the A, P and N domains; sequence motifs of the autoinhibitory domain are colour-coded. **b**, Alignment of the N domains from E2P^{inhib} and adenosine 5′-(β,γ-methylene)triphosphate (AMPPCP)-bound SERCA[36] (Protein Data Bank (PDB) accession code 1T5S), highlighting the overlap of the autoinhibitory domain and the ATP-binding site. Drs2p is coloured as in **a**, SERCA is dark grey and AMPPCP is yellow. **c**, Two-dimensional (2D) class average of E2P^{inhib}; arrow indicates a fuzzy linker between TM10 and the autoinhibitory domain. **d**, Partial release of autoinhibition by PI4P. The map density for the autoinhibitory domain is shown in dark green, with the H1$^{C\text{-}tail}$ highlighted in light green to emphasize its disassociation after the binding of PI4P. The first and last residues that are modelled around the disordered linker are identified. The lipids occupying the regulatory sites in E2P^{inter} and E2P^{active} are identified: phosphatidylserine (black sticks) and PI4P (purple sticks).

Overall structure and conformation

The structures of Drs2p–Cdc50p in the E2P^{inhib}, E2P^{inter} and E2P^{active} conformations were determined at resolutions of 2.8, 3.7 and 2.9 Å, respectively, and reveal functional sites and the PI4P-dependent regulation of Drs2p–Cdc50p (Extended Data Table 1). Complete models, except for minor disordered regions at the termini, were obtained (Supplementary Video 1).

The structure of the Drs2p subunit is typical of P-type ATPases (see Supplementary Discussion) with ten transmembrane helices and three cytosolic domains: the actuator (A) domain, the nucleotide-binding (N) domain and the phosphorylation (P) domain (Fig. 1, Extended Data Fig. 1c, e–g). The Cdc50p subunit has an ectodomain with two asymmetric lobes. The first is dominated by an antiparallel β-sandwich, and the other contains little secondary structure apart from short helical segments (Extended Data Fig. 3a). Two previously identified[27] disulfide bonds are evident (Extended Data Fig. 3b), and at least one *N*-acetylglucosamine unit of each of the four glycosylation sites is revealed by the map (Extended Data Fig. 3c–e). The fold of the ectodomain is similar to the lipid-binding protein seipin, although some loops of Cdc50p are considerably longer than those of seipin (Extended Data Fig. 3f). The two transmembrane helices of Cdc50p extend from the N terminus and the C terminus of the first lobe of the ectodomain (Extended Data Fig. 3a) and interact closely with each other and with transmembrane helix 10 (TM10) of Drs2p (Extended Data Fig. 3g).

Extensive quaternary interactions appear on the luminal face, in which conserved regions of the Cdc50p ectodomain interact with all of the luminal loops of Drs2p (Fig. 1b, Extended Data Fig. 3h–k). The TM3–TM4 loop stretches into a conserved interaction site at the ectodomain of Cdc50p (Extended Data Fig. 3h), and the N terminus of Cdc50p extends along the cytosolic side of the transmembrane domain of Drs2p and makes contacts with a segment (residues 529–538) that connects TM4 and the phosphorylation site of the P domain (Extended Data Fig. 3a, i). This segment is conserved in length in P-type ATPases and couples the chemistry of the phosphorylation site with conformational changes of the transmembrane domain[28], and in P4-ATPases, specifically, the segment is ten residues longer. For a detailed description of Drs2p–Cdc50p interactions, see Supplementary Information.

Autoinhibition and PI4P binding

Our samples reveal three distinct states that lead from autoinhibition to activation. In E2P^{inhib}, the autoinhibitory C terminus forms an extensive interface (residues 1252–1307) that spans the cytosolic domains (Fig. 2a). A short helical segment of the C-terminal tail (H1$^{C\text{-}tail}$; residues 1252–1263) interacts with a unique helical insertion on the P domain, whereas the rest of the tail extends to position a conserved GFAFS motif (residues 1274–1278) at the vertex between the cytosolic domains (Fig. 2a), and overlaps with the nucleotide-binding site (Fig. 2b). Autoinhibition is further stabilized by the clamping of a short loop region in the N domain (residues 698–704) over the GFAFS motif. Cleavage of the C terminus at residue 1290 results in a 10–20-fold increase in enzyme activity compared to wild-type Drs2p–Cdc50p[26], as it removes the bulk of the C-terminal region that interacts with the A domain (as well as unmodelled terminal residues).

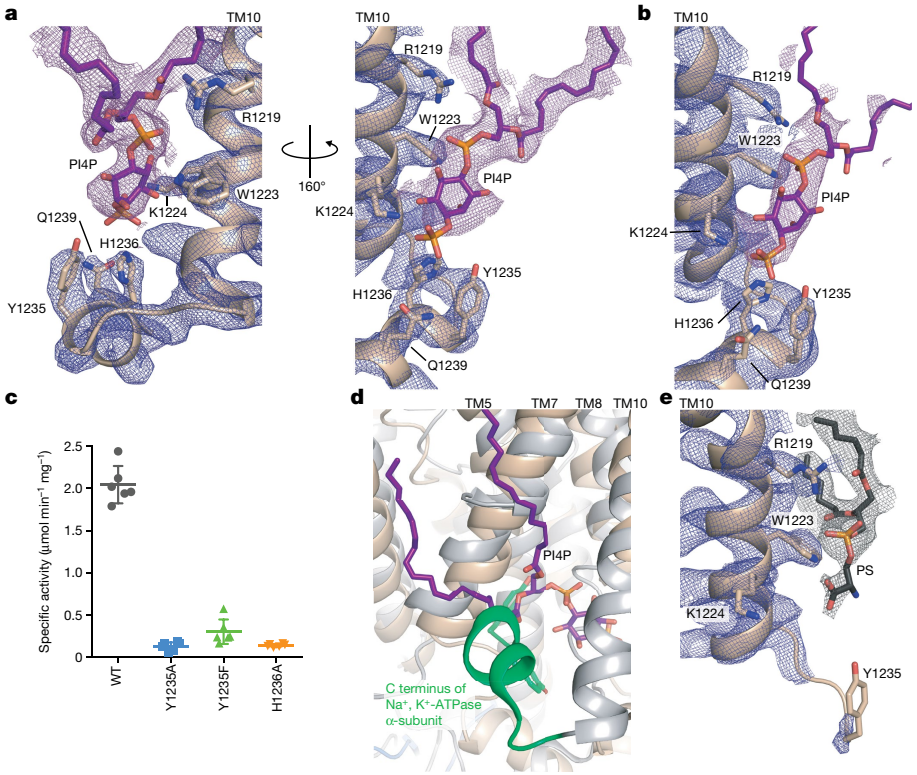


Fig. 3 | Recognition and binding of PI4P by Drs2p. a, The PI4P-binding site of $E2P^{active}$. The cryo-EM map for the lipid (purple) is displayed at a lower threshold (0.75 root mean square deviation (r.m.s.d.)) than for the protein (blue; 2.5 r.m.s.d.). **b**, The PI4P-binding site in $E2P^{inter}$, showing that the PI4P-binding sites in $E2P^{inter}$ and $E2P^{active}$ are consistent. Colours are as in **a**. The cryo-EM maps for lipid and protein are at similar thresholds (1.5 r.m.s.d.). **c**, ATPase activity of wild-type (WT) and three C-terminal mutants of Drs2p–Cdc50p (purified in *n*-dodecyl β-D-maltoside; DDM), plotted as the difference in the rate of ATP hydrolysis observed upon limited proteolysis with trypsin, in the presence of both PI4P and phosphatidylserine, and the rate of ATP hydrolysis observed before the purified protein complexes were added to the assay medium (with phosphatidylserine but in the absence of PI4P; see Extended Data Fig. 4b). Data are mean ± s.d. of six replicates from two independent purification batches. **d**, Superpositioning of Drs2p and the Na^+, K^+-ATPase. Drs2p is shown in tan and the Na^+, K^+-ATPase in grey. The C terminus of the Na^+, K^+-ATPase[30] (PDB 3KDP) is shown in green and overlaps with PI4P. **e**, Phospholipid binding at the PI4P site in $E2P^{inhib}$. Lys1224 moves away and makes no direct contact. Arg1219 makes a non-selective contact with the glycerophosphate group. The cryo-EM map for the lipid (phosphatidylserine (PS); grey) is shown at a lower threshold level (1.0 r.m.s.d.) than the protein (blue, 2.0 r.m.s.d.).

However, truncation at residue 1302 maintains autoinhibition[26] and preserves the interactions with the A domain. This suggests an allosteric mechanism of autoinhibition, in which the cytosolic domains are locked and prevented from undergoing conformational changes. The 16-residue linker between TM10 and the autoregulatory domain is not resolved sufficiently for modelling, but appears at a low-density threshold and in 2D class averages (Fig. 2c).

Density is observed for PI4P bound between TM7, TM8 and TM10 of $E2P^{inter}$ and $E2P^{active}$. Notably, binding of PI4P is concurrent with the movement of the cytosolic ends of TM10 and the transmembrane helices of Cdc50p, and also concurrent with the formation of an amphipathic helix just after TM10. The amphipathic helix propagates away from the autoinhibitory binding site and exerts a pull that dissipates the autoinhibitory interactions of the $H1^{C\text{-}tail}$ with the P domain—thus releasing the P domain. In addition, the amphipathic helix provides a platform for interaction with the TM6–TM7 loop, which moves together with the P domain. This explains in part why C-terminal truncation at residue 1247 preserves PI4P dependence (Extended Data Fig. 1d), whereas truncation at residue 1232 leads to a flippase that is reportedly independent of PI4P[24]. In $E2P^{inter}$, only the downstream interactions of the C terminus with the N and A domains remain preserved (Fig. 2d). The $H1^{C\text{-}tail}$ coincides with the previously described Gea2p-binding site[29], and its release—mediated by PI4P—exposes it for interaction. The PI4P-binding site, however, is notably distinct from a previously proposed site at basic residues 1268–1273[23].

The position of the PI4P glycerol backbone is stabilized by interactions with several positively charged residues that are located in the transmembrane region of Drs2p, and selectivity for PI4P is provided by the interaction of Tyr1235 and His1236 (displayed by the amphipathic helix) with the inositol-4-phosphate group of PI4P (Fig. 3a, b). Mutations of these residues strongly attenuate ATPase activity and reduce the susceptibility of the C-terminal tail to limited proteolysis, which indicates that there is a loss of PI4P binding (Fig. 3c, Extended Data Fig. 4, Supplementary Information). Notably, the PI4P-binding site of Drs2p is located in the same region as the C-terminal YY motif of the α-subunit of the Na^+, K^+-ATPase[30]—a motif that considerably affects the transport function of this enzyme[31](Fig. 3d).

In $E2P^{inhib}$, the amphipathic helix is not present, and the vacant PI4P site contains a lipid with a markedly smaller head group that shows no specific interactions (Fig. 3e). We modelled it as phosphatidylserine—which was the only lipid added to the sample during purification—and presume that regular phospholipids are bound in a non-selective manner in the autoinhibited state.

A putative site for substrate entry

$E2P^{active}$ and $E2P^{inter}$ are largely similar. However, the lack of an autoinhibitory domain in $E2P^{active}$ allows for further rearrangements of the N and A domains, which causes the transmembrane domain to exhibit a conformation that is more open than $E2P^{inter}$ and $E2P^{inhib}$. In this arrangement, movements of the transmembrane regions TM1 and—in particular—TM2 expose the unwound segment of TM4 to the luminal leaflet of the membrane (Fig. 4a, b). This conserved PISL motif has been implicated in lipid transport[14]. The region of TM4 that is exposed to the lumen is lined by TM1, TM2 and TM6, and we propose that

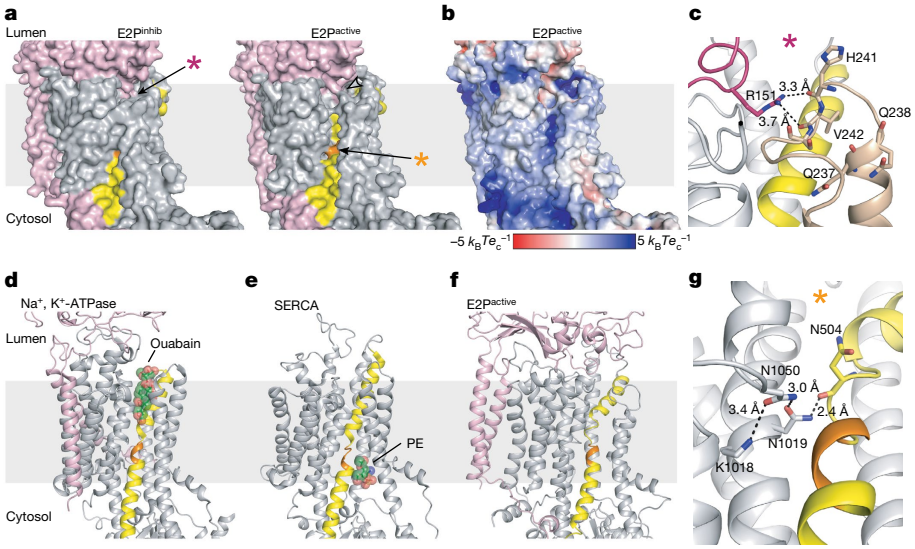


Fig. 4 | Drs2p is activated after the binding of PI4P and release of the autoinhibitory domain. a, A proposed lipid-entry pathway is revealed after full activation (that is, PI4P binding and release of the autoinhibitory domain). TM4 is shown in yellow and the PISL motif is orange. The magenta and orange asterisks mark the locations that are shown in detail in **c** and **g**, respectively. **b**, Electrostatic potential surface of the lipid-entry pathway in E2P^{active}. Electrostatic potential surfaces from APBS[37,38] are shown for all three E2P states in Extended Data Fig. 5a. The colour bar units are dimensionless units of $k_BTe_c^{-1}$, in which k_B is Boltzmann's constant, T is the temperature and e_c is the charge of an electron. **c**, Interaction between Arg151 of Cdc50p and the TM1–TM2 loop of Drs2p in E2P^{inhib}. This location is marked with a magenta asterisk in **a**. Colours are as in **a**; TM1–TM2 is shown in tan. **d**–**f**, Comparison of ouabain-bound Na^+, K^+-ATPase in the E2P state (**d**; PDB 4HYT); SERCA in the E2 state, with phosphatidylethanolamine (PE) bound between TM2 and TM4 (**e**; PDB 2AGV); and Drs2p–Cdc50p in the E2P^{active} state (**f**; this work). View and colours are as in **a**. For the Na^+, K^+-ATPase (**d**), the β- and γ-subunits are pink, TM4 is yellow, the PEGL motif is orange and ouabain is green. For SERCA (**e**), colours are as in **d**, and phosphatidylethanolamine molecules are depicted as green spheres. The ouabain-binding site of the Na^+, K^+-ATPase overlaps with the cleft in E2P^{active}–Drs2p–Cdc50p. The bound phosphatidylethanolamine in SERCA suggests a putative lipid-exit site in Drs2p–Cdc50p. **g**, The side chain of Lys1018, near the PISL motif of TM4. This location is marked with an orange asterisk in **a**.

this cleft (which is empty in our structures) marks the entry point of a lipid-transport pathway. The cleft partially overlaps with a previously proposed entry gate, and contains residues that are important for lipid specificity[12]. In particular, Gln237, which is part of a conserved QQ motif at the end of TM1, points into the open cleft—supporting the role of Gln237 in substrate specificity (Fig. 4c). The cleft is also consistent with a previously proposed hydrophobic gate model[7] in which the conserved isoleucine residue (Ile508 in Drs2p) of the PISL motif[14] has a central role—although the conformation of TM1 and TM2 in E2P^{active} (and the other E2P states reported here) differs from that suggested in previous homology models.

Chimeric constructs and mutants of the TM1–TM2 loop in Dnf1p and Drs2p lipid flippases have indicated that this loop may have a role in lipid binding and specificity[12]. A conserved arginine residue in the ectodomain of Cdc50p (Arg151) reaches towards the proposed entry site—a position from which it could help to orient the luminal loop between TM1 and TM2 and guide the binding of lipids (Fig. 4c). As a comparison, in the Na^+, K^+-ATPase, mutations in the TM1–TM2 loop confer resistance to ouabain[32], and the ouabain-binding site overlaps with the putative lipid-entry pathway in Drs2p–Cdc50p (Fig. 4d, f). Notably, the putative lipid-entry pathway in Drs2p does not span the entire membrane, thus pointing to an alternating-access mechanism of transport. Extension of the pathway towards the cytosolic side overlaps with a lipid-binding site as well as a binding site for the inhibitor cyclopiazonic acid in the sarcoplasmic/endoplasmic reticulum Ca^{2+}-ATPase (SERCA)[33] (Fig. 4e, f). A proposed exit site such as this one, at the cytoplasmic leaflet, would be expected to emerge in a subsequent E2–E1 transition of the functional cycle, and hints at possible evolutionary links between lipid flippases and ion pumps.

Unlike in the cation-transporting P2-ATPases (Extended Data Fig. 5b, c), negatively charged side chains are absent in the transmembrane core of Drs2p, but the potentially positively charged Lys1018 of TM5 interacts with Asn1050 at a bulge of TM6 and could have a stabilizing, yet dynamic, role (Fig. 4g). Lys1018 has previously been implicated in transport in the bovine flippase ATP8A2[7], and it projects a potential positive charge at the middle of the transmembrane pathway (Fig. 4a, b, g). The estimated pK_a value[34] of Lys1018 is around 6.8 for E2P^{inhib} and 7.4 for E2P^{active}, which indicates that it may switch between a neutral and a positively charged state as part of dynamic interactions with a negatively charged lipid head group.

Transport mechanism

The movements of the P and N domains in the transition from the constrained E2P^{inhib} state to the E2P^{inter} and E2P^{active} states (Extended Data Fig. 5d) are mirrored in a concomitant movement of the adjacent TM6–TM7 loop towards the amphipathic helix that forms after PI4P binding (Extended Data Fig. 5e). This suggests that binding of PI4P leads to the movement of the P domain. Of note, the position of the dephosphorylation loop of the A domain remains locally constant with respect to the P domain (Extended Data Fig. 1g). Removal of the C terminus allows for increased mobility of the A domain, which may explain why PI4P alone is not sufficient for full activation of the intact enzyme.

On the basis of our results, we propose the following model for autoregulation and exposure of the lipid-entry site. E2P^{inhib} has a closed transmembrane domain, with the cytosolic domains locked by the autoinhibitory C terminus (Fig. 5, Supplementary Video 2). Binding of PI4P to the transmembrane domain induces the formation of an amphipathic helix on the C-terminal side of TM10. This helix has two effects. First, it causes the remainder of the C terminus to partially unfold, thus destabilizing its interaction with the P domain through the displacement of the H1$^{C\text{-}tail}$ (which contains the putative Gea2p-binding site); and second, it forms an interaction site for the TM6–TM7 loop, which then moves concurrently with the P domain. Together, these two movements shift the P and N domains towards the amphipathic helix. A progressive rotation of the A domain leads to a subtle movement of TM2 and (to a lesser extent) TM1, which results in the E2P^{inter} state. Full displacement of the C terminus leads to the E2P^{active} state, in which a further rotation of the N and A domains drives the

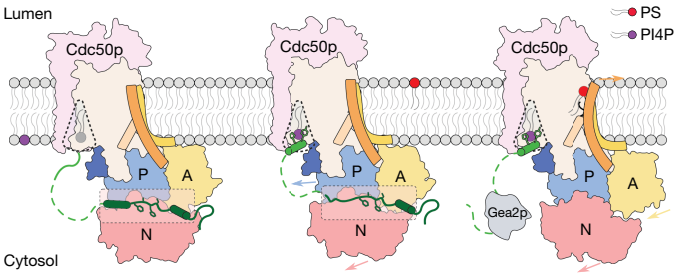


Fig. 5 | Proposed autoregulation mechanism. Schematic of Drs2p–Cdc50p. Domains and regions that are functionally important are coloured: the A, P and N domains of Drs2p are yellow, blue and red, respectively; TM1, TM2 and TM4 are dark yellow, orange and wheat rods, respectively; the TM6–TM7 loop is dark blue; the TM10 amphipathic helix is green; the C terminus is dark green; Cdc50p is pink. Lipids are labelled. The grey surface denotes an auxiliary protein (such as Gea2p) sequestering the C terminus to prevent autoinhibition. Binding of PI4P triggers the formation of the C-terminal amphipathic helix, disrupts the binding of the $H1^{C\text{-}tail}$ to the P domain and results in a rigid-body movement of the cytosolic domains and the TM6–TM7 loop. Full release of the autoinhibitory domain leads to further rearrangements of the N and A domains, movements of TM1 and TM2 and the opening of a putative lipid-entry site.

opening of a putative substrate-binding site through the movement of TM2 away from TM6. We anticipate that subsequent binding of the substrate lipid (phosphatidylserine or phosphatidylethanolamine) will be associated with further conformational changes of the transmembrane domain as dephosphorylation takes place.

Conclusion

The structures of Drs2p–Cdc50p presented here provide insights into the architecture of the P4-ATPase lipid flippases and identify the locations of the Cdc50p subunit and the putative substrate lipid-binding site. The structure of the autoinhibitory domain and the conformational changes associated with the binding of PI4P reveal a mechanism by which P4-ATPases autoregulate, as well as suggesting potential sites at which the activity of lipid flippases could be modulated. Further structures that capture the progressive states of lipid binding, dephosphorylation, lipid translocation and cytoplasmic release may enable the mechanism of the lipid flippase to be determined.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1344-7.

Received: 14 January 2019; Accepted: 28 May 2019;
Published online 26 June 2019.

1. Hankins, H. M., Baldridge, R. D., Xu, P. & Graham, T. R. Role of flippases, scramblases and transfer proteins in phosphatidylserine subcellular distribution. *Traffic* **16**, 35–47 (2015).
2. Montigny, C., Lyons, J., Champeil, P., Nissen, P. & Lenoir, G. On the molecular mechanism of flippase- and scramblase-mediated phospholipid transport. *Biochim. Biophys. Acta* **1861**, 767–783 (2016).
3. Pomorski, T. G. & Menon, A. K. Lipid somersaults: uncovering the mechanisms of protein-mediated lipid flipping. *Prog. Lipid Res.* **64**, 69–84 (2016).
4. Yang, Y., Lee, M. & Fairn, G. D. Phospholipid subcellular localization and dynamics. *J. Biol. Chem.* **293**, 6230–6240 (2018).
5. Ding, J. et al. Identification and functional expression of four isoforms of ATPase II, the putative aminophospholipid translocase. Effect of isoform variation on the ATPase activity and phospholipid specificity. *J. Biol. Chem.* **275**, 23378–23386 (2000).
6. Jacquot, A. et al. Phosphatidylserine stimulation of Drs2p·Cdc50p lipid translocase dephosphorylation is controlled by phosphatidylinositol-4-phosphate. *J. Biol. Chem.* **287**, 13249–13261 (2012).
7. Coleman, J. A., Vestergaard, A. L., Molday, R. S., Vilsen, B. & Andersen, J. P. Critical role of a transmembrane lysine in aminophospholipid transport by mammalian photoreceptor P4-ATPase ATP8A2. *Proc. Natl Acad. Sci. USA* **109**, 1449–1454 (2012).
8. Lee, J. Y. et al. Crystal structure of the human sterol transporter ABCG5/ABCG8. *Nature* **533**, 561–564 (2016).
9. Perez, C. et al. Structure and mechanism of an active lipid-linked oligosaccharide flippase. *Nature* **524**, 433–438 (2015).
10. Brunner, J. D., Lim, N. K., Schenck, S., Duerst, A. & Dutzler, R. X-ray structure of a calcium-activated TMEM16 lipid scramblase. *Nature* **516**, 207–212 (2014).
11. Morra, G. et al. Mechanisms of lipid scrambling by the G protein-coupled receptor opsin. *Structure* **26**, 356–367 (2018).
12. Baldridge, R. D. & Graham, T. R. Two-gate mechanism for phospholipid selection and transport by type IV P-type ATPases. *Proc. Natl Acad. Sci. USA* **110**, E358–E367 (2013).
13. Jensen, M. S. et al. Phospholipid flipping involves a central cavity in P4 ATPases. *Sci. Rep.* **7**, 17621 (2017).
14. Vestergaard, A. L. et al. Critical roles of isoleucine-364 and adjacent residues in a hydrophobic gate control of phospholipid transport by the mammalian P4-ATPase ATP8A2. *Proc. Natl Acad. Sci. USA* **111**, E1334–E1343 (2014).
15. Bryde, S. et al. CDC50 proteins are critical components of the human class-1 P4-ATPase transport machinery. *J. Biol. Chem.* **285**, 40562–40572 (2010).
16. Coleman, J. A. & Molday, R. S. Critical role of the β-subunit CDC50A in the stable expression, assembly, subcellular localization, and lipid transport activity of the P4-ATPase ATP8A2. *J. Biol. Chem.* **286**, 17205–17216 (2011).
17. van der Mark, V. A., Elferink, R. P. & Paulusma, C. C. P4 ATPases: flippases in health and disease. *Int. J. Mol. Sci.* **14**, 7897–7922 (2013).
18. Alder-Baerens, N., Lisman, Q., Luong, L., Pomorski, T. & Holthuis, J. C. Loss of P4 ATPases Drs2p and Dnf3p disrupts aminophospholipid transport and asymmetry in yeast post-Golgi secretory vesicles. *Mol. Biol. Cell* **17**, 1632–1642 (2006).
19. Natarajan, P., Wang, J., Hua, Z. & Graham, T. R. Drs2p-coupled aminophospholipid translocase activity in yeast Golgi membranes and relationship to *in vivo* function. *Proc. Natl Acad. Sci. USA* **101**, 10614–10619 (2004).
20. Azouaoui, H. et al. A high-yield co-expression system for the purification of an intact Drs2p–Cdc50p lipid flippase complex, critically dependent on and stabilized by phosphatidylinositol-4-phosphate. *PLoS ONE* **9**, e112176 (2014).
21. Zhou, X. & Graham, T. R. Reconstitution of phospholipid translocase activity with purified Drs2p, a type-IV P-type ATPase from budding yeast. *Proc. Natl Acad. Sci. USA* **106**, 16586–16591 (2009).
22. Xu, P., Baldridge, R. D., Chi, R. J., Burd, C. G. & Graham, T. R. Phosphatidylserine flipping enhances membrane curvature and negative charge required for vesicular transport. *J. Cell Biol.* **202**, 875–886 (2013).
23. Natarajan, P. et al. Regulation of a Golgi flippase by phosphoinositides and an ArfGEF. *Nat. Cell Biol.* **11**, 1421–1426 (2009).
24. Zhou, X., Sebastian, T. T. & Graham, T. R. Auto-inhibition of Drs2p, a yeast phospholipid flippase, by its carboxyl-terminal tail. *J. Biol. Chem.* **288**, 31807–31815 (2013).
25. Tsai, P. C., Hsu, J. W., Liu, Y. W., Chen, K. Y. & Lee, F. J. Arl1p regulates spatial membrane organization at the trans-Golgi network through interaction with Arf-GEF Gea2p and flippase Drs2p. *Proc. Natl Acad. Sci. USA* **110**, E668–E677 (2013).
26. Azouaoui, H. et al. High phosphatidylinositol 4-phosphate (PI4P)-dependent ATPase activity for the Drs2p–Cdc50p flippase after removal of its N- and C-terminal extensions. *J. Biol. Chem.* **292**, 7954–7970 (2017).
27. Puts, C. F. et al. Mapping functional interactions in a heterodimeric phospholipid pump. *J. Biol. Chem.* **287**, 30529–30540 (2012).
28. Møller, J. V., Olesen, C., Winther, A. M. & Nissen, P. The sarcoplasmic Ca^{2+}-ATPase: design of a perfect chemi-osmotic pump. *Q. Rev. Biophys.* **43**, 501–566 (2010).
29. Chantalat, S. et al. The Arf activator Gea2p and the P-type ATPase Drs2p interact at the Golgi in *Saccharomyces cerevisiae*. *J. Cell Sci.* **117**, 711–722 (2004).
30. Morth, J. P. et al. Crystal structure of the sodium–potassium pump. *Nature* **450**, 1043–1049 (2007).
31. Poulsen, H. et al. Neurological disease mutations compromise a C-terminal ion pathway in the Na^+/K^+-ATPase. *Nature* **467**, 99–102 (2010).
32. Price, E. M. & Lingrel, J. B. Structure–function relationships in the Na,K-ATPase α subunit: site-directed mutagenesis of glutamine-111 to arginine and asparagine-122 to aspartic acid generates a ouabain-resistant enzyme. *Biochemistry* **27**, 8400–8408 (1988).
33. Obara, K. et al. Structural role of countertransport revealed in Ca^{2+} pump crystal structure in the absence of Ca^{2+}. *Proc. Natl Acad. Sci. USA* **102**, 14489–14496 (2005).
34. Olsson, M. H., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: consistent treatment of internal and surface residues in empirical pK_a predictions. *J. Chem. Theory Comput.* **7**, 525–537 (2011).
35. Jakobi, A. J., Wilmanns, M. & Sachse, C. Model-based local density sharpening of cryo-EM maps. *eLife* **6**, e27131 (2017).
36. Sørensen, T. L., Møller, J. V. & Nissen, P. Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science* **304**, 1672–1675 (2004).
37. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: an automated pipeline for the setup of Poisson–Boltzmann electrostatics calculations. *Nucleic Acids Res.* **32**, W665–W667 (2004).
38. Jurrus, E. et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **27**, 112–128 (2018).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Biochemical studies of C-terminal Drs2p–Cdc50p mutants. Expression and streptavidin purification of wild-type and mutant forms of Drs2p–Cdc50p for functional studies were carried out as previously described[20]. Specifically, we used for this purpose a C-terminal tobacco etch virus (TEV)-cleavable BAD tag, and DDM was used throughout the purification procedure. Size-exclusion chromatography (SEC) was performed on a Superdex 200 Increase 10/300GL column, with a mobile phase containing 0.5 mg ml^{-1} DDM and 0.025 mg ml^{-1} POPS in 50 mM MOPS-Tris buffer at pH 7.0, containing 100 mM KCl, 20% glycerol (w/v) and 5 mM $MgCl_2$ (SSR). ATP hydrolysis by PI4P-binding mutants was measured at 30 °C using an enzyme-coupled assay, by continuously monitoring the rate of NADH oxidation at 340 nm[39]. The purified Drs2p–Cdc50p complexes were added at about 2 μg ml^{-1}, in a cuvette containing SSR buffer supplemented with 1 mM ATP, 1 mM phosphoenolpyruvate, 0.4 mg ml^{-1} pyruvate kinase, 0.1 mg ml^{-1} lactate dehydrogenase, 0.25 mM NADH, 1 mM NaN_3, 1 mg ml^{-1} DDM and 0.1 mg ml^{-1} POPS. Trypsin and PI4P were subsequently added to concentrations of 0.05 mg ml^{-1} and 0.025 mg ml^{-1}, respectively. Conversion of NADH oxidation rates that were expressed in AU s^{-1} to ATPase activities in μmol min^{-1} mg^{-1} was based on the extinction coefficient of NADH (around 6,200 M^{-1} cm^{-1}) and on quantification of Drs2p by SDS–PAGE stained with Coomassie blue, using known amounts of SEC-purified Drs2p as standards.

For limited proteolysis experiments[26], wild-type and mutant forms of Drs2p–Cdc50p were purified on streptavidin beads, either in the presence or in the absence of POPS. Proteolysis with 40 U ml^{-1} thrombin took place for 1 h at 20 °C in the presence of 1 mg ml^{-1} DDM, and 0.05 mg ml^{-1} phosphatidylserine and/or 0.025 mg ml^{-1} PI4P when needed. Proteolysis was quenched by adding 1 mM PMSF.

Expression and purification of Drs2p–Cdc50p for structural studies. Protein expression in *S. cerevisiae*, membrane collection and solubilization were performed as previously described[20,26,40] using a C-terminal thrombin-cleavable BAD-tagged Drs2p construct that produces Drs2p$^{\Delta N104}$–Cdc50p, in which Drs2p is missing the first 104 residues owing to thrombin cleavage. A second construct with an additional thrombin site after residue 1247 in Drs2p resulted in the N- and C-terminally truncated Drs2p$^{\Delta N104/C1247}$–Cdc50p sample.

Affinity chromatography on streptavidin resin and detergent exchange. The BAD-tagged protein was batch-bound to free streptavidin sepharose resin (typically 1 ml resin per 60 ml of solubilized material) for 1 h at 4 °C. Detergent exchange into LMNG was performed, by washing with 2 column volumes (CV) SSR with 1 mM dithiothreitol (DTT) and 0.2 mg ml^{-1} LMNG, followed by washing with 10 CV SSR with 1 mM DTT and 0.1 mg ml^{-1} LMNG. The resin was resuspended in 1 CV SSR with 1 mM DTT, 0.1 mg ml^{-1} LMNG and 50 μg ml^{-1} brain phosphatidylserine (Avanti Polar Lipids). Bovine thrombin (4 units per ml resin; Calbiochem) was added to cleave the protein off the resin during an overnight incubation at 4 °C. The protein was eluted from the resin in 10–20 CV SSR with 1 mM DTT and 0.1 mg ml^{-1} LMNG, and concentrated to 0.5–1 ml in a 100-kDa centrifugal concentrator (Vivaspin) with the sample typically reaching concentrations of 5–10 mg ml^{-1}.

Cleavage of double-truncated construct. To produce the double-truncated Drs2p$^{\Delta N104/C1247}$–Cdc50p protein, after elution from the resin and concentration, 4 U bovine thrombin per ml resin used for the purification was added, along with 0.025 mg ml^{-1} brain PI4P (Avanti Polar Lipids). This was followed by incubation at room temperature for 1 h and quenching of the protease activity with 1 mM PMSF.

SEC. For Drs2p$^{\Delta N104}$–Cdc50p, SEC was run on a Superdex 200 Increase 10/300 column on an ÄKTA purifier system at 4 °C in SSR with 0% glycerol, 1 mM DTT and 0.03 mg ml^{-1} LMNG. The peak fractions typically resulted in a Drs2p–Cdc50p concentration of 0.6 mg ml^{-1}, which was used directly for preparation of cryo-EM grids, or stored at −80 °C for later use. For Drs2p$^{\Delta N104/\Delta C1247}$–Cdc50p, a first round of SEC was run on a TSKg4000SW silica column on an ÄKTA purifier system at 4 °C in SSR with 0% glycerol, 1 mM DTT and 0.03 mg ml^{-1} LMNG. The peak fractions were pooled and concentrated to 8 mg ml^{-1} using a centrifugal concentrator with a cut-off of 50 kDa, and stored at −80 °C for later use. A second round of SEC was run on an analytical Superdex 200 Increase 3.2/300 column on an ÄKTA purifier system at 4 °C in SSR with 0% glycerol, 1 mM DTT and 0.03 mg ml^{-1} LMNG, in which a 50-μl sample was injected, to remove the background detergent produced by concentrating the sample. Pooling of the peak fractions resulted in a protein concentration of 0.6 mg ml^{-1}, which was used directly for preparation of cryo-EM grids. Representative chromatograms and gels are shown in Extended Data Fig. 2a–c.

Activity measurement on purified protein for structural studies. The activity of the purified Drs2p–Cdc50p used for structural studies was assayed using an arsenic-based Baginski assay[41], which is a colorimetric assay for free inorganic phosphate. Drs2p–Cdc50p in LMNG to a final concentration of 10 μg ml^{-1} was added to a reaction buffer of SSR with 0% glycerol, 1 mM DTT, 0.02 mg ml^{-1} LMNG and 5 mM NaN_3 (final concentrations). Phosphatidylserine (with 8-carbon acyl chains (C8:0)), brain PI4P and BeF_3^- ($BeSO_4$ and KF in a 1:20 molar ratio) were added, when present, to final concentrations of 78 μg ml^{-1}, 20 μg ml^{-1} and 1 mM, respectively. After addition of protein to the reaction buffers, the samples were incubated on ice for 1 h, before transfer to 30 °C, and after reaching this temperature, the reactions were initiated by addition of ATP to a concentration of 4 mM. At specific time points, 50 μl of the sample was transferred to a 96-well microplate, and mixed with 50 μl 1:5 solution of 30 mM ammonium heptamolybdate in H_2O, and 0.17 M ascorbic acid and 0.1% SDS in 0.5 M HCl. After 10 min at room temperature, 75 μl arsenic solution (2% (w/v) anhydrous sodium metaarsenic, 2% (w/v) trisodium citrate dehydrate and 2% (v/v) glacial acid) was added to prevent further complexing of molybdate by phosphate. The plate was left at room temperature for 30 min, before measurement of the absorbance at 860 nm on a Wallac Victor 3 Multilabel Plate Reader (Perkin Elmer).

Negative-stain electron microscopy. Copper G400-C3 grids were coated with 2% celluidine, followed by evaporation of amorphous carbon using a Leica EM SCD500 high vacuum sputter coater. Before use, the grids were glow-discharged on a PELCO easiGlow Glow Discharge Cleaning System at 25 mA for 45 s. A total of 3 μl protein sample diluted to 20 μg ml^{-1} in detergent-free buffer was added, followed by staining 3 times with 3 μl 2% uranyl formate solution, which had been stored at −80 °C. Micrographs were collected on a Tecnai G2 Spirit (120 kV) with a Tietz F416 CCD camera using Leginon[42]. Imaging was performed at 67,000× magnification with a binned camera (pixel size 3.15 Å). Data processing including contrast transfer function (CTF) estimation, particle picking, extraction using a box size of 84 pixels and 2D classification was performed in cisTEM[43] (Extended Data Fig. 2d–f).

Grid preparation for cryo-EM. C-flat Holey Carbon grids, CF-1.2/1.3-4C (Protochips), were glow-discharged on a PELCO easiGlow Glow Discharge Cleaning System at 15 mA for 45 s before addition of 3 μl of 0.6 mg ml^{-1} Drs2p–Cdc50p in LMNG, which had been incubated on ice for at least 1 h with 1 mM $BeSO_4$, 20 mM KF, and 0.1 mg ml^{-1} brain PI4P when indicated. The samples were vitrified on a Vitrobot IV (Thermo Fisher) at 4 °C and 100% humidity.

Cryo-EM data collection. The data were acquired on a Titan Krios with an X-FEG operated at 300 kV. Movies were acquired using a Gatan K2 camera with a Bioquantum energy filter operated at a slit width of 20 eV. Movies were collected in counting mode with a calibrated pixel size of 1.077 Å per pixel at a magnification of 130,000× (Max Planck Institute for Biophysics). Exposures of 8 s fractionated into 40 frames were collected through EPU software (Thermo Fisher) at a dose rate of 1.4 or 1.5 e$^-$ per Å^2 per frame, corresponding to a total dose of 56 or 60 e$^-$ per Å^2.

For Drs2p$^{\Delta N104}$–Cdc50p, 765 movies were collected on samples with 0.1 mg ml^{-1} PI4P and 3,069 movies were collected without PI4P. For Drs2p$^{\Delta N104/C1247}$–Cdc50p, 2,391 movies were collected on a grid with 0.1 mg ml^{-1} PI4P and 801 movies were collected on a grid without additional PI4P. However, after initial processing the datasets resulted in identical reconstructions with the same density in the PI4P-binding site (probably because of the PI4P that was added during the purification of this sample), and they were treated as one dataset going forward.

Cryo-EM data processing. For all three datasets, movie alignment with dose weighting using all frames and CTF determination was performed in cisTEM through Unblur[44] and CTFFIND4[45], respectively. After manual inspection of the micrographs, 2,050 were selected for Drs2p$^{\Delta N104/C1247}$–Cdc50p with PI4P, 2,687 for Drs2p$^{\Delta N104}$–Cdc50p and 701 for Drs2p$^{\Delta N104}$–Cdc50p with PI4P. Using the cisTEM reference-free particle picker, a total of 1,047,615 particles were picked for Drs2p$^{\Delta N104/C1247}$–Cdc50p with PI4P, 2,156,578 for Drs2p$^{\Delta N104}$–Cdc50p and 578,440 for Drs2p$^{\Delta N104}$–Cdc50p with PI4P. The particles were extracted in cisTEM using a box size of 256 pixels, and cisTEM was used for 2D classification (although this was not used for selecting good particles).

For Drs2p$^{\Delta N104}$–Cdc50p (E2P^{inhib}), three ab initio 3D references were generated in cryoSPARC[46] from all particles, which resulted in one class that corresponded to the protein particle and two that corresponded to junk. Three rounds of heterogeneous 3D classification in cryoSPARC were performed in which the first round had one protein class and four junk classes, whereas the next rounds only used two junk classes and one protein class. This resulted in 769,469 particles, which were subjected to heterogeneous 3D refinement, resulting in an initial reconstruction at 3.2 Å from 752,881 particles. These particles were re-extracted in RELION-3[47] from movies that were aligned through MotionCor2[48] (implemented in RELION). Per-particle CTF refinement, with estimation of the beam tilt and Bayesian polishing, was performed in RELION-3[49], before the final 3D refinement that resulted in an unmasked resolution of 3.0 Å and a masked resolution of 2.8 Å. RELION-3 was used for estimation of the local resolution. The processing strategy is summarized in Extended Data Fig. 6.

For Drs2p$^{\Delta N104}$–Cdc50p with PI4P (E2P^{inter}), three ab initio 3D references were generated in cryoSPARC[46], resulting in one protein-like class and two junk classes. These were used as reference in 3 rounds of heterogeneous 3D classification that resulted in 291,944 protein particles. An initial homogeneous 3D refinement of these particles resulted in a reconstruction at 3.8 Å from 277,569 particles. These particles were re-extracted in RELION-3[47] from movies that were aligned through

MotionCor2[48] (implemented in RELION). Per-particle CTF refinement, with estimation of the beam tilt and Bayesian polishing, was performed in RELION-3[49], before the final 3D refinement that resulted in an unmasked resolution of 3.4 Å. However, this appeared to be a mixed state near TM10 of Drs2p, the PI4P-binding site and in parts of the autoinhibitory domain. As simple 3D classification using the refined map as a reference did not separate this heterogeneity, five new ab initio references were generated in cryoSPARC[46] with similarity 1.0, to allow for classes that were very similar. These were then all used as references for heterogeneous 3D refinement, resulting in one class that corresponded to a PI4P-bound structure, two classes identical to the determined autoinhibited structure in the absence of PI4P and two minor junk classes. The PI4P-bound class resulted in a reconstruction at 3.7 Å from 78,981 particles from cryoSPARC; this reconstruction was also refined in RELION-3 to an unmasked resolution of 3.9 Å and a masked resolution of 3.7 Å after processing. RELION-3 was used for estimation of the local resolution. The processing strategy is summarized in Extended Data Fig. 7.

For Drs2p$^{\Delta N104/C1247}$–Cdc50p with PI4P (E2^{active}), three ab initio 3D references were generated in cryoSPARC[46] from all particles, resulting in one protein-like class and two junk classes. All particles were then subjected to heterogeneous 3D classification in cryoSPARC, using each junk reference twice and the protein-like class once, which resulted in four junk classes and one class with particles that corresponded to the protein. This class was then subjected to heterogeneous 3D refinement in cryoSPARC, resulting in an initial reconstruction at 3.4 Å from 493,753 particles. The 3D refinement was repeated in RELION-2.1[50] using a soft solvent mask, which resulted in a reconstruction of similar quality. To improve the map, particle sorting was performed in RELION-2.1; particles with a z-score above 0.9 or a defocus higher than 2.0 μm were rejected, and the remaining 418,512 were re-extracted in RELION-3[47] from movies that were aligned through MotionCor2[48] (implemented in RELION). Per-particle CTF refinement, with estimation of the beam tilt and Bayesian polishing, was performed in RELION-3[49], before the final 3D refinement that resulted in an unmasked resolution of 3.1 Å and a masked resolution of 2.9 Å. RELION-3 was used for estimation of the local resolution. The processing strategy is summarized in Extended Data Fig. 8.

Data collection and processing statistics are summarized in Extended Data Table 1.

Model building and refinement. The Drs2p$^{\Delta N104/C1247}$–Cdc50p with PI4P was built manually in Coot[51], guided by secondary structure predictions from RaptorX[52] and—for Drs2p—by the structures of the Na^+, K^+-ATPase and SERCA in the E2P state (PDB 4HYT and 3B9B, respectively), owing to the shared topology and similarity of the E2P conformations in these proteins and Drs2p. For Cdc50p, the positions of all glycosylation sites were visible as at least one sugar moiety, and these were used to validate the de novo traced structure along with the presence of the two disulfide bonds. The model was refined using Namdinator[53] and phenix.real_space_refine[54].

Drs2p$^{\Delta N104}$–Cdc50p was built by fitting the Drs2p$^{\Delta N104/C1247}$–Cdc50p model using Namdinator[53], followed by manual editing in Coot and manual tracing of the autoregulatory C terminus. Coordinate refinement was carried out using phenix.real_space_refine[54]. The Drs2p$^{\Delta N104}$–Cdc50p with PI4P model was built based on the other two structures, with manual editing in Coot[51], followed by phenix.real_space_refine[54].

For Drs2p, the 78 N-terminal residues of the construct, as well as the 46 C-terminal residues and the linker region of 16 residues (20 residues in the structure of E2P^{inter}) between TM10 and the autoregulatory domain in Drs2p$^{\Delta N104}$, were too disordered for modelling or were entirely missing from the density. For Cdc50p, the 19 N-terminal residues (18 in the structure of E2P^{active}) and the entire C-terminal tail of 33 residues (35 in the structure of E2P^{inhib}) could not be modelled.

Model validation was performed using MolProbity[55] in PHENIX[56].

Modelling and refinement statistics are summarized in Extended Data Table 1. Model-to-map Fourier shell correlation (FSC) curves and representative densities from different areas of the maps are shown in Extended Data Fig. 9.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Cryo-EM maps for the *S. cerevisiae* Drs2p–Cdc50p in the E2P^{inhib}, E2P^{inter} and E2P^{active} forms are available on the Electron Microscopy Data Bank under accession numbers EMD-4972 (E2P^{inhib}; that is, E2–BeF_3^-), EMD-4973 (E2P^{inter}; that is, E2–BeF_3^-–PI4P) and EMD-4974 (E2P^{active}; that is, C-terminally truncated E2–BeF_3^-–PI4P). Coordinates of the atomic structures have been deposited in the PDB under accession numbers 6ROH, 6ROI and 6ROJ.

39. Møller, J. V., Lind, K. E. & Andersen, J. P. Enzyme kinetics and substrate stabilization of detergent-solubilized and membraneous (Ca^{2+} + Mg^{2+})-activated ATPase from sarcoplasmic reticulum. Effect of protein–protein interactions. *J. Biol. Chem.* **255**, 1912–1920 (1980).
40. Azouaoui, H. et al. Coordinated overexpression in yeast of a P4-ATPase and its associated Cdc50 subunit: the case of the Drs2p/Cdc50p lipid flippase complex. *Methods Mol. Biol.* **1377**, 37–55 (2016).
41. Baginski, E. S., Foa, P. P. & Zak, B. Determination of phosphate: study of labile organic phosphate interference. *Clin. Chim. Acta* **15**, 155–158 (1967).
42. Suloway, C. et al. Automated molecular microscopy: the new Leginon system. *J. Struct. Biol.* **151**, 41–60 (2005).
43. Grant, T., Rohou, A. & Grigorieff, N. *cis*TEM, user-friendly software for single-particle image processing. *eLife* **7**, e35383 (2018).
44. Grant, T. & Grigorieff, N. Automatic estimation and correction of anisotropic magnification distortion in electron microscopes. *J. Struct. Biol.* **192**, 204–208 (2015).
45. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
46. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
47. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *eLife* **7**, e42166 (2018).
48. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
49. Zivanov, J., Nakane, T. & Scheres, S. H. W. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. *IUCrJ* **6**, 5–17 (2019).
50. Fernandez-Leiro, R. & Scheres, S. H. W. A pipeline approach to single-particle processing in *RELION*. *Acta Crystallogr. D* **73**, 496–502 (2017).
51. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *Coot*. *Acta Crystallogr. D* **66**, 486–501 (2010).
52. Wang, S., Li, W., Liu, S. & Xu, J. RaptorX-Property: a web server for protein structure property prediction. *Nucleic Acids Res.* **44**, W430–435 (2016).
53. Kidmose, R. T. et al. Namdinator - automatic Molecular Dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ.* **6**, https://doi.org/10.1107/S2052252519007619 (2019).
54. Afonine, P. V. et al. Real-space refinement in *PHENIX* for cryo-EM and crystallography. *Acta Crystallogr. D* **74**, 531–544 (2018).
55. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
56. Adams, P. D. et al. *PHENIX*: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
57. Yan, R. et al. Human SEIPIN binds anionic phospholipids. *Dev. Cell* **47**, 248–256 (2018).
58. Ashkenazy, H. et al. ConSurf 2016: an improved methodology to estimate and visualize evolutionary conservation in macromolecules. *Nucleic Acids Res.* **44**, W344–W350 (2016).
59. Sievers, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).
60. Afonine, P. V. et al. New tools for the analysis and validation of cryo-EM maps and atomic models. *Acta Crystallogr. D* **74**, 814–840 (2018).

Acknowledgements We thank A. M. Nielsen and T. Klymchuk for technical assistance; P. Gourdon and C. Grønberg for early contributions on samples; P. Champeil for initial functional characterization of purified Drs2p–Cdc50p and for critically reading the manuscript; D. Mills and colleagues at the Max Planck Institute for Biophysics (Frankfurt/Main) and R. Liebrechts at the iNANO center (Aarhus University) for support with data collection and discussions. This work was supported by grants from the Danish National Research Foundation for the PUMPkin center of excellence and from the Lundbeck Foundation for the Brainstruc center of excellence (2015-2666) to P.N.; by an EMBO Long-Term Fellowship to M.-R.A.; by postdoctoral grants from the Danish Council for Independent Research (0602-02912B) and the Lundbeck Foundation (R171-2014-663 and R209-2015-2704) to J.A.L.; by a PhD fellowship from the Boehringer-Ingelheim Fonds to M.T.; by an ANR grant (ANR-14-CE09-0022), the French Infrastructure for Integrated Structural Biology (FRISBI; ANR-10-INSB-05) and the Centre National de la Recherche Scientifique (CNRS) to G.L.; and by the German Research Foundation to A.M. (Mo2752/2).

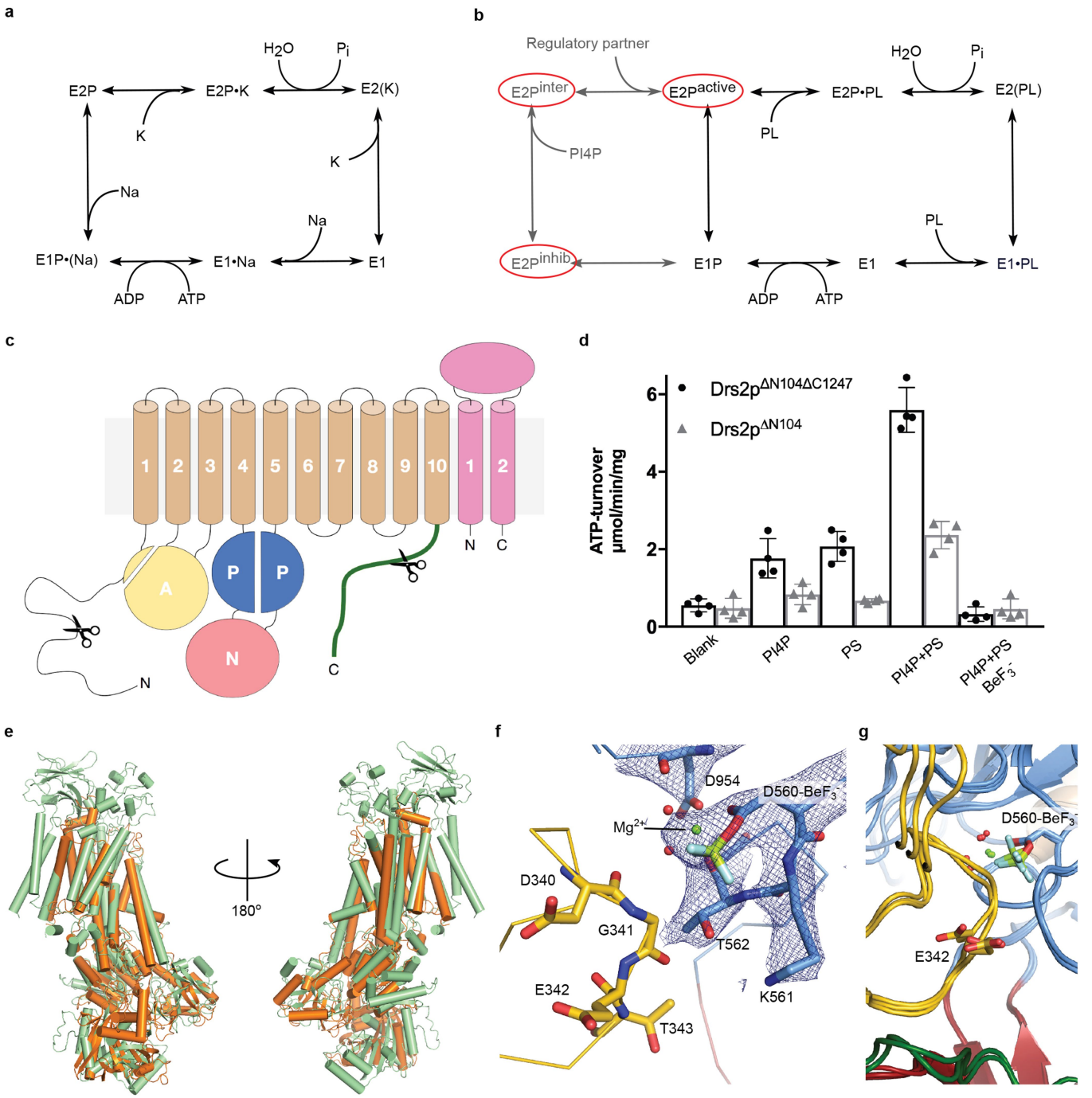
Author contributions P.N. and G.L. conceived the project, and J.A.L., T.B. and P.N. defined the cryo-EM study with A.M. and W.K. The samples were characterized and developed by M.T., J.J.U., J.A.L. and M.-R.A., together with T.D., C.M. and G.L., and exploratory electron microscopy studies were performed by J.A.L., M.T., J.J.U. and T.B. Cryo-EM analysis was performed by M.T., D.J., J.A.L., T.B. and A.M. Data processing and 3D reconstruction were performed by M.T., with support and advice from D.J., J.A.L., J.L.K. and A.M. Model building and refinement were performed by M.T. and J.A.L., with assistance from J.J.U. and J.L.K. Mutant forms were prepared and functionally characterized by T.D., C.M. and G.L. P.N. and J.A.L. supervised the project together with A.M. The manuscript was drafted by M.T., J.A.L. and P.N. All authors commented on the manuscript.

Competing interests The authors declare no competing interests.

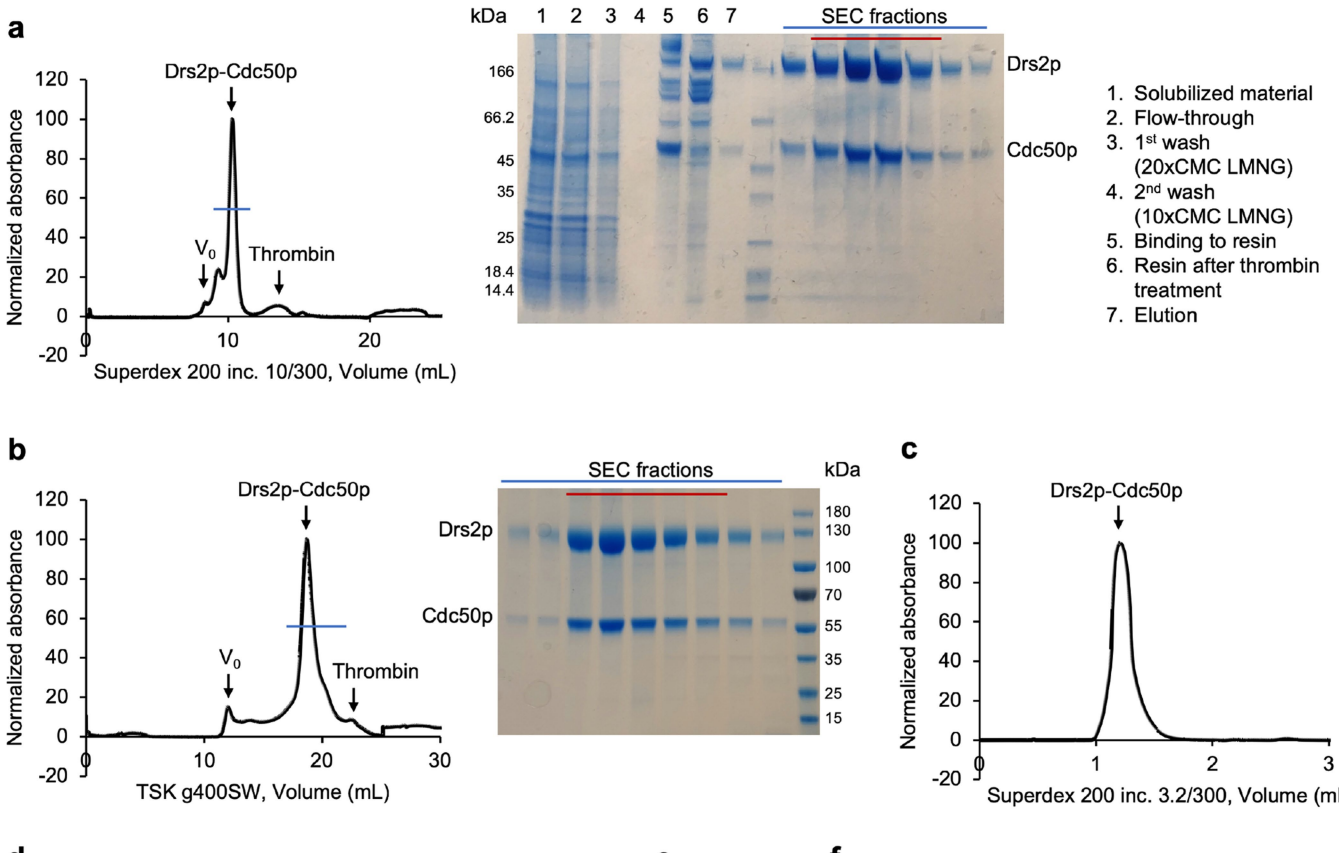
Additional information

Correspondence and requests for materials should be addressed to G.L., A.M. or P.N.

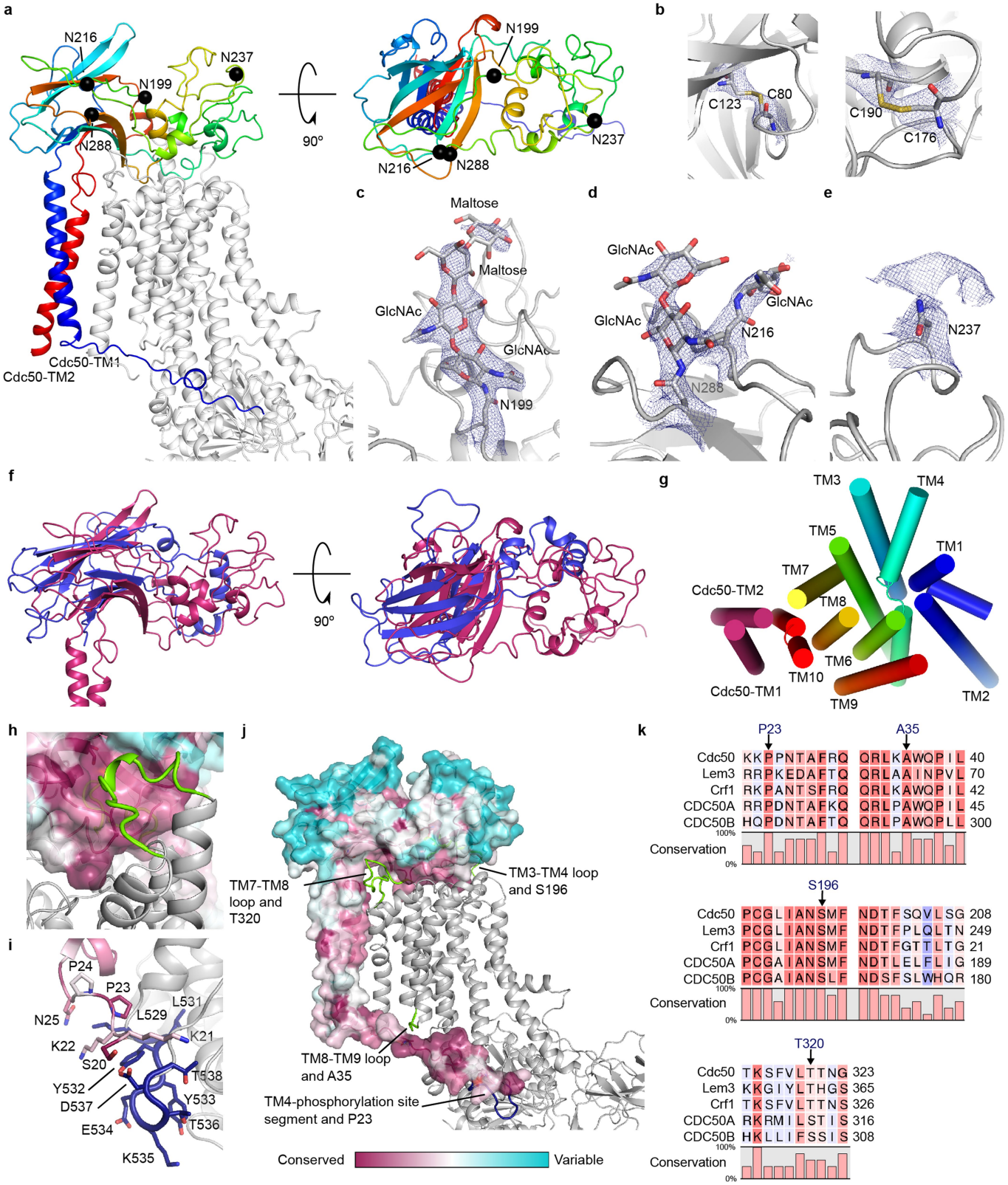
Peer review information *Nature* thanks Joost Holthuis and the other anonymous reviewer(s) for their contribution to the peer review of this work.



Extended Data Fig. 1 | Activity of samples used for structural studies. **a**, Post–Albers cycle for the Na^+, K^+-ATPase. Sodium is transported during the E1 half-cycle, and potassium during the E2 half-cycle. P_i, inorganic phosphate. **b**, Post–Albers cycle for Drs2p–Cdc50p, indicating the off-cycle regulation (grey). Lipid is transported during the E2 half-cycle. The structures that were determined in this work are marked with red circles. PL, translocated phospholipid. **c**, Topology of Drs2p–Cdc50p, indicating the cleavages at the termini of Drs2p of the constructs that were used for structural studies (ΔN104: all constructs; ΔC1247: $E2P^{active}$). Cdc50p is pink; for Drs2p the transmembrane domain is tan, the A, P and N domains are yellow, blue and red, respectively, and the autoinhibitory C terminus is green. **d**, Specific activity of $Drs2p^{\Delta N104/C1247}$–Cdc50p and $Drs2p^{\Delta N104}$–Cdc50p in LMNG, measured by Baginski assay. Water-soluble phosphatidylserine (C8:0), brain PI4P and BeF_3^- were added as indicated to final concentrations of 78 μg ml^{-1}, 20 μg ml^{-1} and 5 mM, respectively. Data are mean ± s.d. of four replicates from two independent purification batches. **e**, Alignment of Drs2p–Cdc50p (in the $E2P^{active}$ form) and SERCA (PDB 3B9B) (superpositioning of the C_α carbons, excluding the N domain). Drs2p–Cdc50p is green; SERCA is orange. **f**, The phosphorylation site of $E2P^{active}$, showing density for the BeF_3^- inhibitor, Mg^{2+} ion and coordinating residues. The characteristic E2P conformation of the dephosphorylation loop, in which the glutamate points away from the phosphorylation site, is shown in stick representation. **g**, The three Drs2p–Cdc50p structures aligned based on the P domain. Asp560–BeF_3^- and Glu342 are shown to illustrate the similar conformations. Colours are as in Fig. 1a.

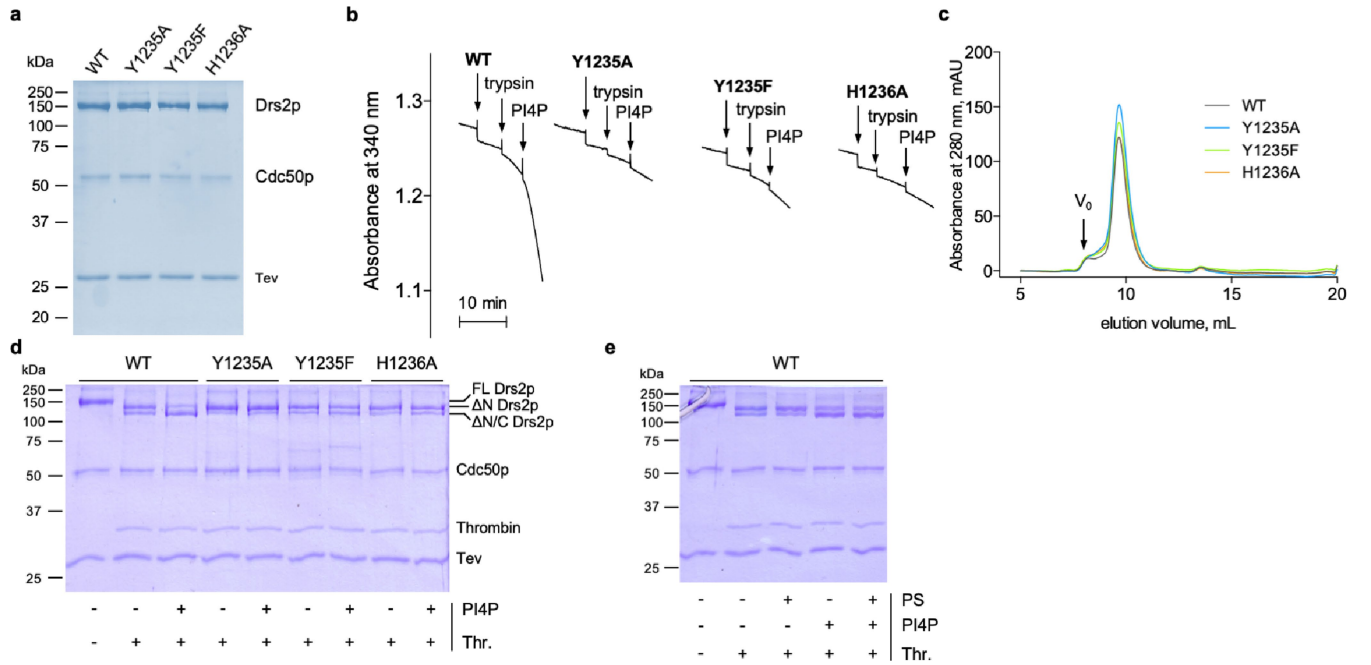


Extended Data Fig. 2 | Purification and negative-stain electron microscopy of samples for structural studies. a, SEC chromatogram and gel of Drs2p$^{\Delta N104}$–Cdc50p. The red line on the SDS–PAGE image indicates the pooled fractions. **b**, SEC chromatogram and gel of Drs2p$^{\Delta N104/C1247}$–Cdc50p; gel shows the SEC fractions. Red line as in **a**. **c**, Analytical SEC of Drs2p$^{\Delta N104/C1247}$–Cdc50p purified from **b**; this was used for cryo-EM. **d**, A representative negative-stained micrograph of autoinhibited Drs2p$^{\Delta 104}$–Cdc50p in LMNG. **e**, Representative 2D class averages of the sample in **d** show well-defined and homogeneous particles with recognizable features of P-type ATPases. **f**, Enlarged 2D class average, highlighting the recognizable domains of Drs2p–Cdc50p.

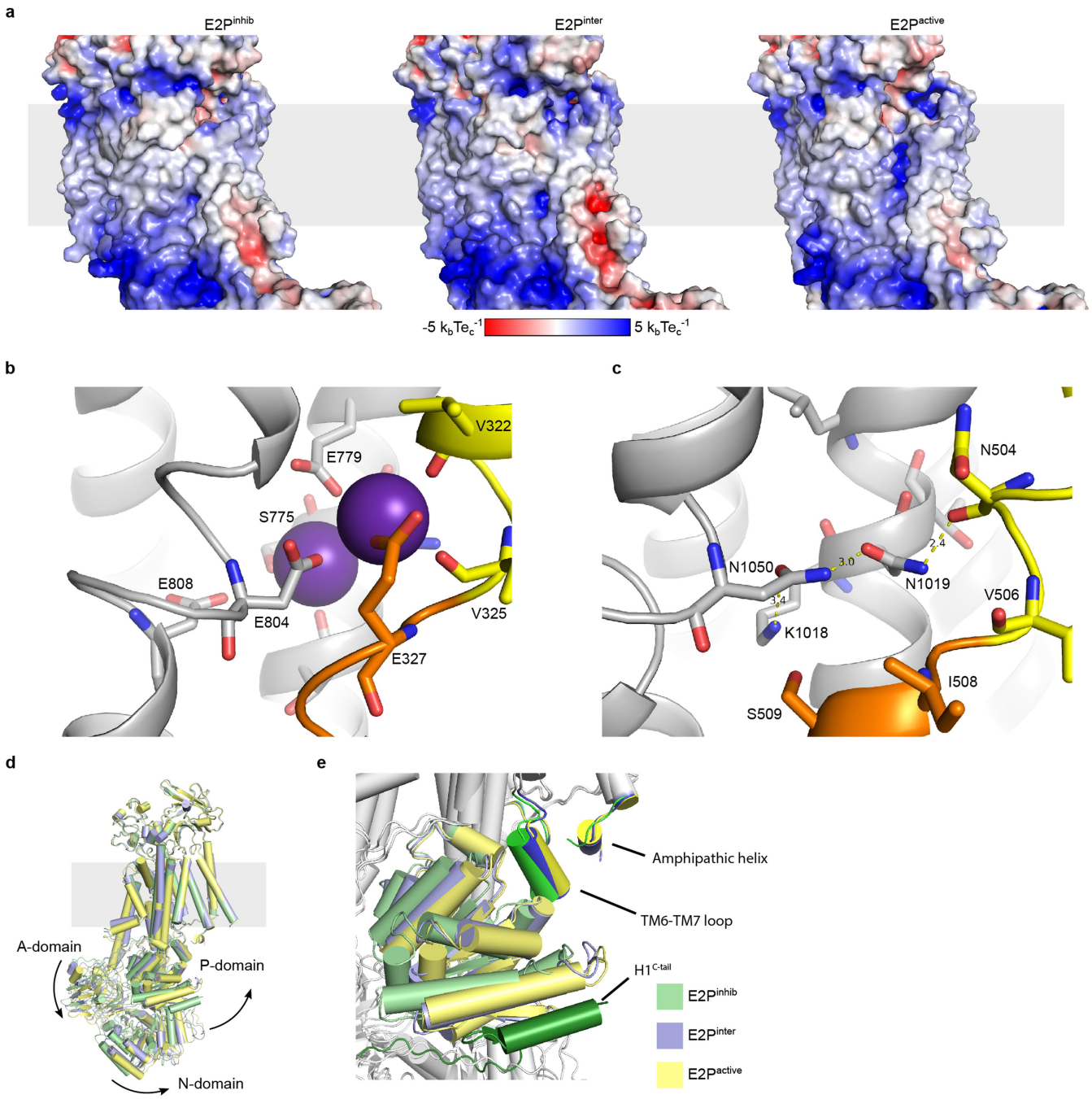


Extended Data Fig. 3 | See next page for caption.

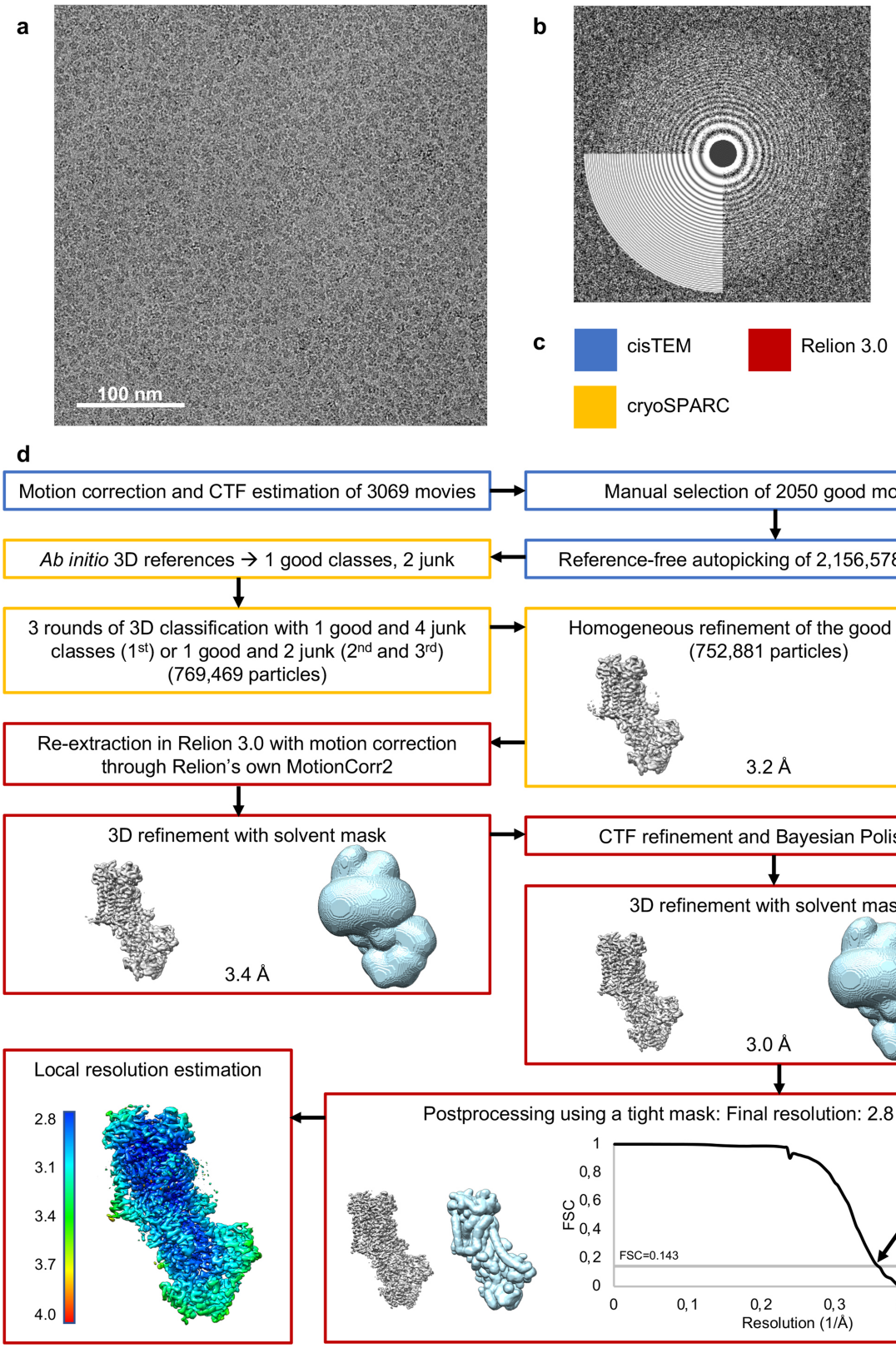
Extended Data Fig. 3 | Fold of Cdc50p. a, Cdc50p from $E2P^{inhib}$, shown in rainbow colours. C_α carbons of glycosylated asparagine residues are shown as black spheres; Drs2p is shown in grey. **b**, Position of the disulfide sites in Cdc50p from $E2P^{inhib}$. **c**–**e**, Glycosylation sites in Cdc50p from $E2P^{inhib}$. Map levels are 1.75 r.m.s.d. in **c** and **d**. The density for the glycosylation at Asn237 is only apparent at lower map levels (0.75 r.m.s.d.) and has not been modelled. **f**, Alignment of Cdc50p (magenta cartoon) and a monomer of human seipin[57], a lipid-binding protein (blue cartoon, PDB 6DS5), illustrating the similar folds (although some loops of Cdc50p are more extensive). The sequence identity between the two proteins is only 4%. Transmembrane helices of seipin that extend from similar positions to the helices of Cdc50p are not present in the structure, but may extend into the membrane in a similar manner to that observed for Cdc50p. **g**, The transmembrane helices of the Drs2p–Cdc50p complex viewed from the luminal side. The transmembrane helices of Drs2p are rainbow coloured; Cdc50p transmembrane helices are pink. **h**, Interaction between the Cdc50p ectodomain (shown as surface) and the luminal TM3–TM4 loop of Drs2p (light green). **i**, Interaction between the N terminus of Cdc50p and the segment of Drs2p that leads from TM4 to the phosphorylation site. Residues 529–538 are not present in P2-ATPases and are shown in dark blue. **j**, Segments of Drs2p that were found to interact with Cdc50p mutants that disrupt the formation of the complex are highlighted in green, and the insert in Drs2p between TM4 and the phosphorylation site is blue. The structure shown in **h**–**j** is $E2P^{inhib}$. Cdc50p is coloured on the basis of the conservation using ConSurf[58]. **k**, Part of a sequence alignment of proteins of the CDC50 family from *S. cerevisiae* (Cdc50p, Lem3p and Crf1p) and human CDC50A and CDC50B. Residues that are important for complex formation are identified. Uniprot identifiers are as follows: Cdc50p, P25656; Lem3p, P42838; Crf1p, P53740; CDC50A, Q9NV96; CDC50B, Q3MIR4. The alignment was performed using Clustal Omega[59]. For full sequence alignment, see Supplementary Fig. 1.



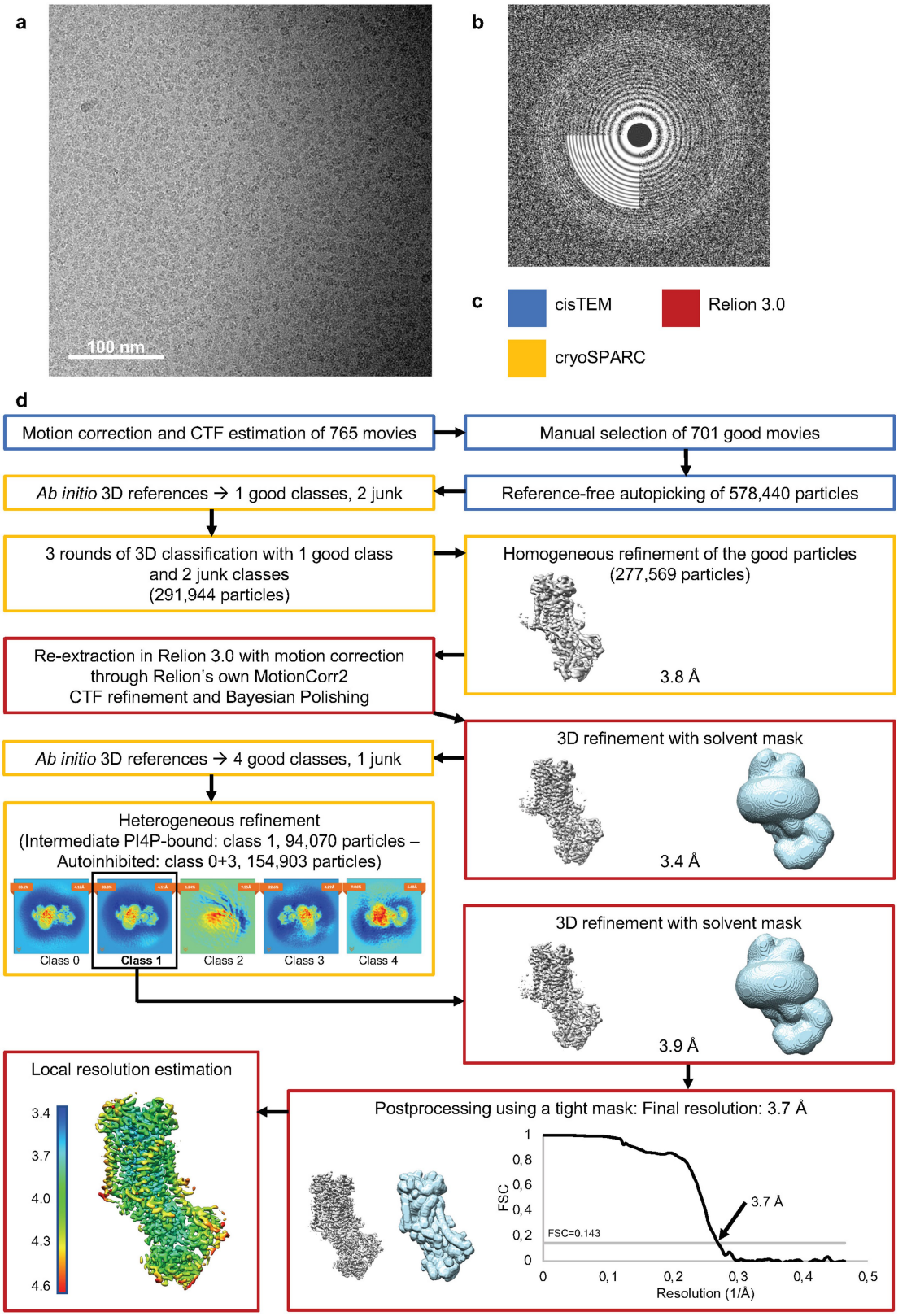
Extended Data Fig. 4 | Characterization of PI4P-binding mutants. **a**, Coomassie-blue-stained SDS–PAGE of streptavidin-purified wild-type and mutant Drs2p–Cdc50p. Tobacco etch virus protease (TEV) was used to release the complex from the streptavidin beads. **b**, ATPase activity of PI4P-binding mutants, using an enzyme-coupled assay. The assay medium contained 1 mM ATP, 0.1 mg ml^{-1} POPS and 1 mg ml^{-1} DDM in SSR buffer. The rate of ATP hydrolysis was continuously recorded at 340 nm after subsequent addition of 2 μg ml^{-1} of the purified complex, 5 μg ml^{-1} trypsin and 25 μg ml^{-1} PI4P. **c**, SEC on a Superdex 200 10/300GL column. The arrow indicates the dead volume (V_0). **d**, Limited proteolysis of streptavidin-purified wild-type and mutant Drs2p–Cdc50p by thrombin (Thr.). Control sample in lane 1 contains full-length (FL) Drs2p–Cdc50p. In the absence of PI4P, wild-type and mutant forms of Drs2p are truncated to sizes that correspond to the loss of the first 104 N-terminal residues (ΔN Drs2p; top band in lanes 2, 4, 6 and 8) and to the loss of both the first 104 N-terminal residues and the last 65 C-terminal residues (ΔN/C Drs2p; second band in lanes 2, 4, 6 and 8). Incubation with PI4P during limited proteolysis (lanes 3, 5, 7 and 9) promotes further C-terminal truncation only for the wild-type Drs2p, indicating that the Drs2p mutants are less sensitive to PI4P. **e**, Control proteolysis experiment showing that phosphatidylserine does not alter the cleavage of wild-type Drs2p–Cdc50p with thrombin, and thereby demonstrating that the C-terminal mutants Y1235A, Y1235F and H1236A are specifically defective in PI4P binding.



Extended Data Fig. 5 | Proposed lipid-translocation pathway and comparison between the three Drs2p–Cdc50p structures. a, A proposed pathway for the translocation of lipids is revealed after activation by PI4P binding and C-terminal truncation. Electrostatic surfaces were calculated in APBS[37,38]. See also Fig. 4a. **b**, Binding of potassium in the Na^+, K^+-ATPase[30] (PDB 3KDP), showing the coordination of the ions by negatively charged and polar residues. TM4 is coloured yellow, the PEGL motif is orange and the bound potassium ions are purple spheres. **c**, Sites and residues of $E2P^{active}$ that correspond to the ion-binding sites of the Na^+, K^+-ATPase in **b**. TM4 is yellow, the PISL motif is orange and stabilizing hydrogen bonds are shown. **d**, Superpositioning of the three Drs2p–Cdc50p structures, based on Cdc50p and TM7–TM10 of Drs2p. $E2P^{inhib}$ is green, $E2P^{inter}$ is blue and $E2P^{active}$ is yellow. The major conformational changes are in the cytosolic domains, TM1 and TM2. **e**, Expanded view of **d**, highlighting the changes in the P domains of the three Drs2p–Cdc50p structures. The TM6–TM7 loop and the $H1^{C\text{-}tail}$ are depicted in slightly darker colours and the autoinhibitory domain of $E2P^{inhib}$ is dark green. Destabilization or removal of the helical segment of the autoinhibitory domain ($E2P^{inter}$ and $E2P^{active}$) leads to a rigid-body movement of the P domain.

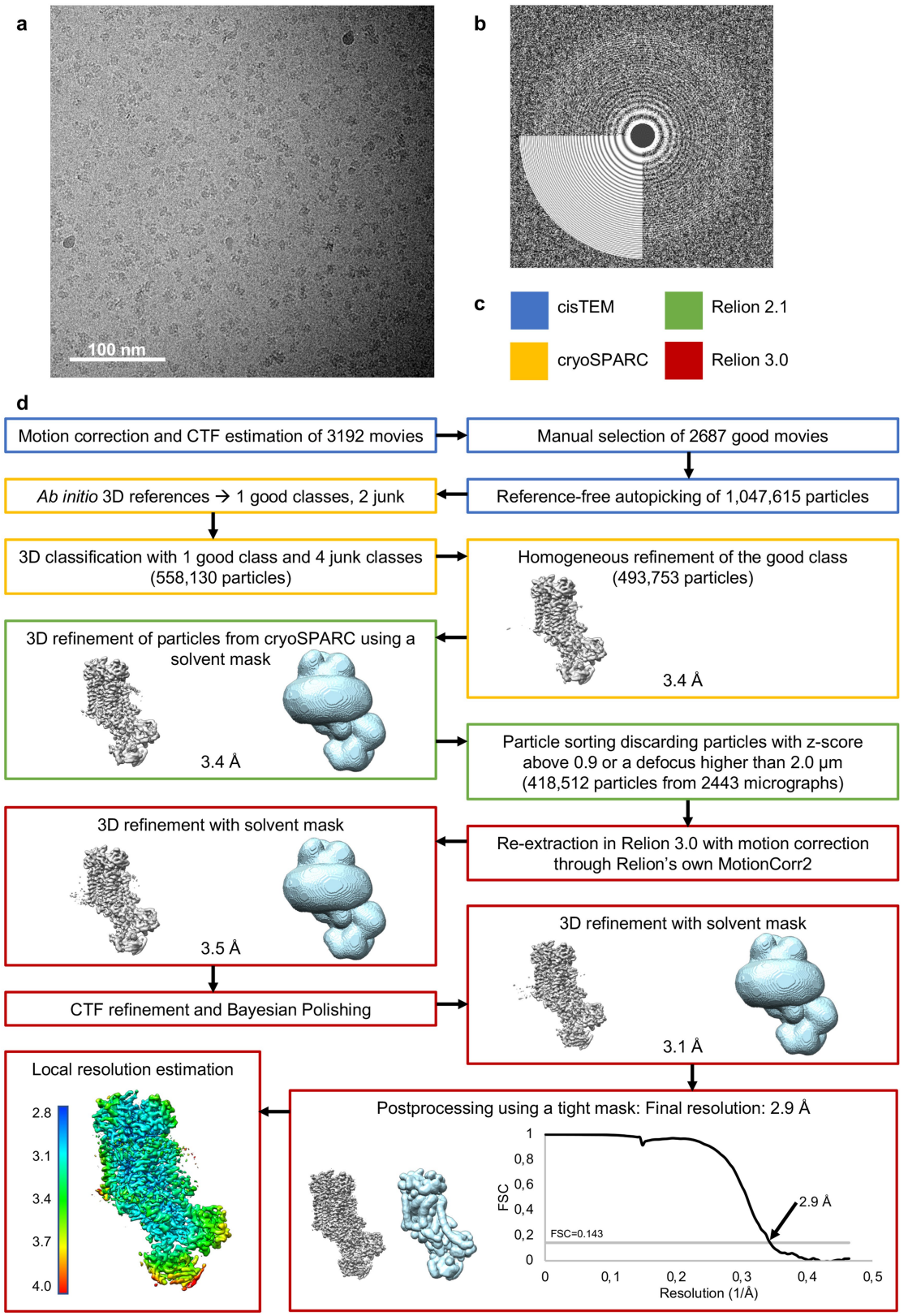


Extended Data Fig. 6 | Processing pipeline for cryo-EM data of Drs2p–Cdc50p E2P^{inhib}. **a**, Representative motion-corrected and dose-weighted micrograph (defocus 1.6 μm) of autoinhibited Drs2p$^{\Delta N104}$–Cdc50p in LMNG, frozen at a concentration of 0.6 mg ml^{-1}. **b**, Fourier power spectrum of the micrograph shown in **a**, as well as the fit from CTFFIND 4.1 through cisTEM, which extends to 3 Å. **c**, Colour code for the processing software. **d**, Data-processing workflow, indicating the number of particles that remained after each step at which particles were discarded. The densities resulting from 3D refinement are shown in grey, and relevant masks are light blue. The resolutions listed for 3D refinements are at FSC = 0.143.

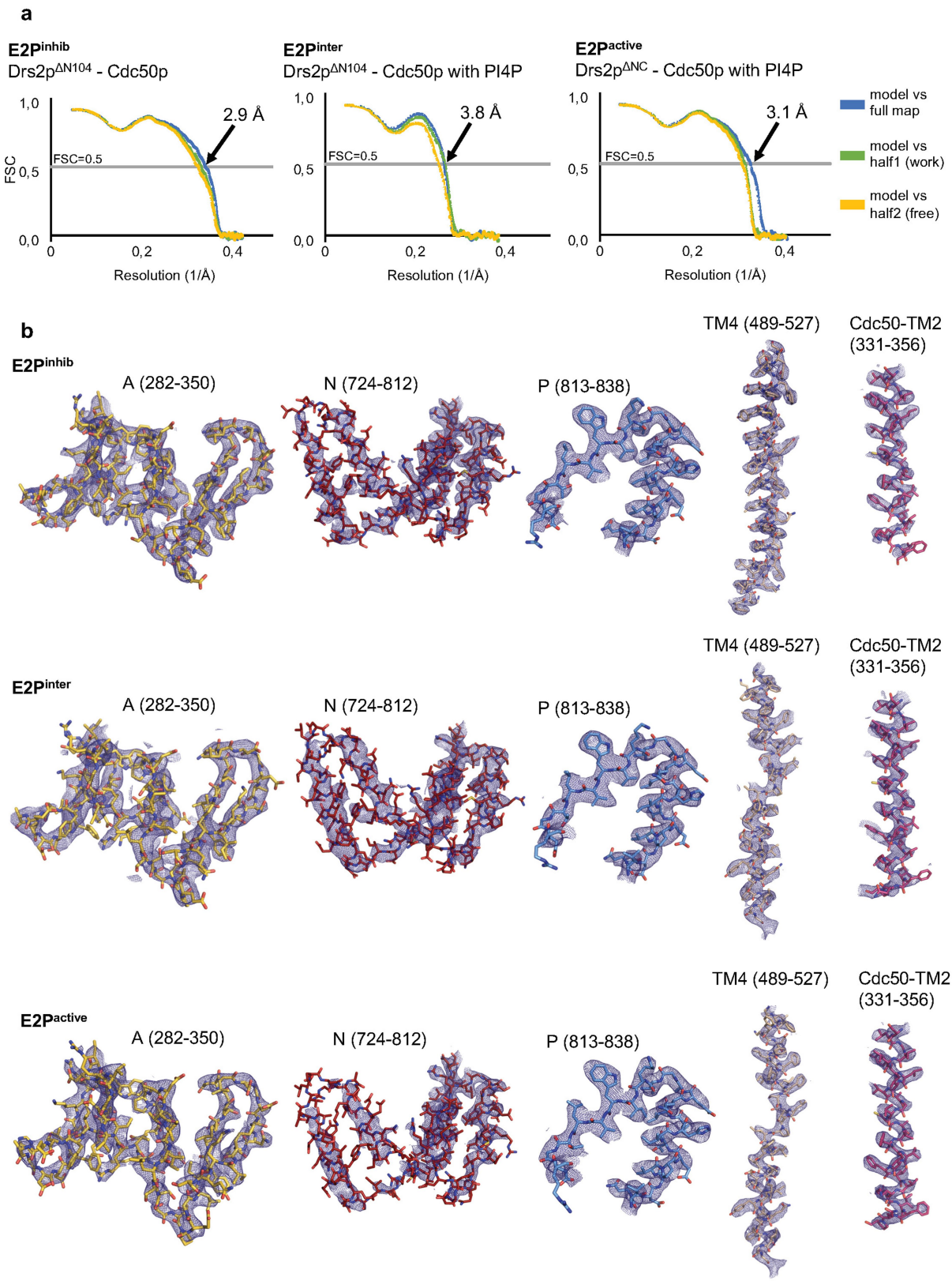


Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Processing pipeline for cryo-EM data of Drs2p–Cdc50p E2P^{inter}. **a**, Representative motion-corrected and dose-weighted micrograph (defocus 1.5 μm) of Drs2p$^{\Delta N104}$–Cdc50p with an intact C terminus in LMNG, frozen at a concentration of 0.6 mg ml^{-1} in the presence of 75 μg ml^{-1} brain PI4P. **b**, Fourier power spectrum of micrograph in **a**, as well as the fit from CTFFIND 4.1 through cisTEM, which extends to 5 Å. **c**, Colour code for the processing software. **d**, Data-processing workflow, indicating the number of particles that remained after each step at which particles were discarded. The densities resulting from 3D refinement are shown in grey, and relevant masks are light blue. The resolutions listed for 3D refinements are at FSC = 0.143. The 3.4-Å refinement suggested a mixed state around TM10. To classify the structural heterogeneity that was caused by incomplete binding of PI4P, new ab initio references were generated in cryoSPARC, allowing for high similarity because the conformations were expected to be similar. Two different conformations resulted: the autoinhibited one and a PI4P-bound version. The autoinhibited conformation was identical to E2P^{inhib} and adding these particles to the E2P^{inhib} dataset did not improve the reconstruction. The PI4P-bound conformation was further refined in RELION.



Extended Data Fig. 8 | Processing pipeline for cryo-EM data of Drs2p–Cdc50p E2P^{active}. **a**, Representative motion-corrected and dose weighted micrograph (defocus of 1.7 µm) of C-terminally truncated Drs2p$^{\Delta NC}$–Cdc50p in LMNG, frozen at a concentration of 0.6 mg ml^{-1} in the presence of 75 µg ml^{-1} brain PI4P. **b**, Fourier power spectrum of the micrograph shown in **a**, as well as the fit from CTFFIND 4.1 through cisTEM, which extends to 3 Å. **c**, Colour code for the processing software. **d**, Data-processing workflow, indicating the number of particles that remained after each step at which particles were discarded. The densities resulting from 3D refinement are shown in grey, and relevant masks are light blue. The resolutions listed for 3D refinements are at FSC = 0.143.



Extended Data Fig. 9 | Model validation and representative densities. **a**, Cross-validation FSC curves for map-to-model fit produced by Mtriage[60]. Curves representing the model versus the full map are calculated based on the final model and the full, filtered and sharpened map that it was refined against. For the model versus half-maps, the model (before the final refinement) was refined against half-map 1 (filtered and sharpened as for the full map), and FSC curves were calculated using this refined model against each half-map. **b**, Representative densities from different areas of the three LocScale maps. Each segment is labelled with the residues that are shown, and demonstrates the quality of the map in specific areas. All densities are shown at a threshold of 1.5 r.m.s.d.

Extended Data Table 1 | Cryo-EM data collection, refinement and validation statistics

	Drs2p ΔN104/Cdc50p (autoinhibited) (EMD-4972) (PDB 6ROH)	Drs2p ΔN104/Cdc50p with PI4P (intermediate) (EMD-4973) (PDB 6ROI)	Drs2p ΔNC/Cdc50p with PI4P (activated) (EMDB-4974) (PDB 6ROJ)
Data collection and processing			
Magnification	130,000	130,000	130,000
Voltage (kV)	300	300	300
Microscope	Titan Krios	Titan Krios	Titan Krios
Electron exposure (e–/Å^2)	60	60	56
Defocus range (μm)	0.3-3.0 (0.4-2.5)*	0.4-2.4 (0.5-2.4)*	0.3-3.0 (0.4-2.0)*
Pixel size (Å)	1.077	1.077	1.077
Camera	Gatan K2	Gatan K2	Gatan K2
Number of movies	3069 (2050)*	765 (701)*	3192 (2443)*
Symmetry imposed	C1	C1	C1
Initial particle images (no.)	2,157,578	578,440	1,047,615
Final particle images (no.)	752,881	78,981	418,512
Map resolution (Å)	2.8	3.7	2.9
FSC threshold	0.143	0.143	0.143
Map resolution range (Å)	2.7-3.4	3.6-4.4	2.9-4.1
Refinement			
Model resolution (Å)	2.9	3.8	3.1
FSC threshold	0.5	0.5	0.5
Model resolution range (Å)			
Map sharpening *B* factor (Å^2)[#]	-40 (-76)	-60 (-115)	-35 (-69)
Model composition			
Non-hydrogen atoms	11758	11775	11465
Protein residues	1449	1447	1404
Ligands	2 H_2O, 1 Mg^{2+}, 1 PS, 5 NAG, 2 BMA	2 H_2O, 1 Mg^{2+}, 1 PI4P, 5 NAG, 2 BMA	2 H_2O, 1 Mg^{2+}, 1 PI4P, 6 NAG, 4 BMA
B factors (Å^2)			
Protein	33.75/150.55/73.76	35.45/170.28/89.45	34.18/187.73/77.30
Ligand	55.17/92.52/75.64	68.87/120.20/89.63	51.25/111.21/76.81
Water	55.49/55.73/55.61	77.61/80.06/78.84	60.71/61.39/61.05
R.m.s. deviations			
Bond lengths (Å)	0.006	0.008	0.006
Bond angles (°)	0.867	0.992	0.902
Validation			
MolProbity score	1.40	1.61	1.36
Clashscore	4.00	5.05	3.58
Poor rotamers (%)	0.23	0.23	0.16
EMRinger	3.80	2.13	3.50
Ramachandran plot			
Favored (%)	96.60	94.99	96.64
Allowed (%)	3.40	5.01	3.36
Disallowed (%)	0.0	0.0	0.0

*The number in parentheses corresponds to the final range or number.
#The *B*-factor listed was used for refinement, whereas the one in parentheses was estimated by RELION post-processing.

nature research

Corresponding author(s): *Double-blind peer review submissions: write DBPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	EPU (FEI/ThermoFisher) was used for automated collection of cryo-EM data
Data analysis	Cryo-EM data was analyzed through the cisTEM 1.0.0-beta package (including Unblur and CTFFIND 4.1 procedures), Relion 2.1 and 3.0 and cryoSPARC v1. Modelling was performed using COOT 0.8.9.1, secondary structure prediction from RaptorX and Namdinator. Phenix 1.14-3260 was used for refinement, Molprobidity, EMRinger and Mtriage were used for validation. LocScale was used for scaling of maps for visualization. All of the software is published and referenced in the supplementary. All but cryoSPARC are open source. No custom algorithms were used.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

PDB entry codes (6ROH, 6ROI, and 6ROH) and EMBD entry codes (EMD-4972, EMD-4973, and EMD-4974) have been included

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For Drs2 mutants (Figure 3), activity was measured from 3 different samples (1 trace out of the 3 measurements is displayed for each species in extended data Figure 4). We consider this sample size is sufficient because: 1/ activity measurement is based on continous monitoring of NADH oxidation, resulting in accurate measurement of the rate of ATP hydrolysis, as compared with end-point measurements and 2/ the 3 values obtained were very similar to each other
Data exclusions	No data were excluded to plot results displayed in any tables or Figures
Replication	All attempts were included as replicates to plot data displayed in Figure 3 and the reproducibility between the various measurements is illustrated by error bars (standard deviation) and scatter plots. All experiments were reproducible.
Randomization	No group allocation was performed in this study
Blinding	Investigators were not blinded to group allocation as the present study is not a clinical research trial

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Saccharomyces cerevisiae expression strain W303.1b/GAL4 (a, leu2-3, his3-11, trp1-1::TRP1-GAL10-GAL4, ura3-1, ade2-1, canr, cir+)
Authentication	n.a.
Mycoplasma contamination	n.a.
Commonly misidentified lines (See ICLAC register)	n.a.

LETTER

https://doi.org/10.1038/s41586-019-1381-2

A two-qubit gate between phosphorus donor electrons in silicon

Y. He[1,2], S. K. Gorman[1,2], D. Keith[1], L. Kranz[1], J. G. Keizer[1] & M. Y. Simmons[1]*

Electron spin qubits formed by atoms in silicon have large (tens of millielectronvolts) orbital energies and weak spin–orbit coupling, giving rise to isolated electron spin ground states with coherence times of seconds[1,2]. High-fidelity (more than 99.9 per cent) coherent control of such qubits has been demonstrated[3], promising an attractive platform for quantum computing. However, inter-qubit coupling—which is essential for realizing large-scale circuits in atom-based qubits—has not yet been achieved. Exchange interactions between electron spins[4,5] promise fast (gigahertz) gate operations with two-qubit gates, as recently demonstrated in gate-defined silicon quantum dots[6–10]. However, creating a tunable exchange interaction between two electrons bound to phosphorus atom qubits has not been possible until now. This is because it is difficult to determine the atomic distance required to turn the exchange interaction on and off while aligning the atomic circuitry for high-fidelity, independent spin readout. Here we report a fast (about 800 picoseconds) $\sqrt{\text{SWAP}}$ two-qubit exchange gate between phosphorus donor electron spin qubits in silicon using independent single-shot spin readout with a readout fidelity of about 94 per cent on a complete set of basis states. By engineering qubit placement on the atomic scale, we provide a route to the realization and efficient characterization of multi-qubit quantum circuits based on donor qubits in silicon.

Early proposals for spin qubits in semiconductors, such as those by Kane[1] and Loss & DiVincenzo[5], envisaged using large exchange interactions J to couple neighbouring electron and/or nuclear spin qubits. The exchange interaction generates a natural two-qubit gate by conditionally swapping the electron spin states if they are anti-parallel. Experimentally this requires tuning the exchange energy by many orders of magnitude up to gigahertz frequencies over a relatively small range in voltage detuning, making the realization of a SWAP gate challenging. Besides the $\sqrt{\text{SWAP}}$ gate, two other approaches to realizing a two-qubit exchange gate in gate-defined quantum dots have been pursued: the controlled phase (CZ)[11] and the controlled rotation (CROT)[12] operations. When combined with single-qubit rotations, any of these three entangling gates are sufficient to perform universal quantum computation. A quality unique to the SWAP gate is its capability to move quantum information through large arrays of qubits that are only coupled to their neighbours by performing multiple SWAP operations[4]. In particular, donor-based systems proposed by Kane[1], with their tight-confinement potential, minimal gate density[13] and strong capacitive coupling, have the potential of achieving strong exchange coupling[14] for fast SWAP gates in quantum computing applications. However, because of their small Bohr radius, it is difficult to determine how far apart to place the qubits to tune the exchange interaction high enough to perform a $\sqrt{\text{SWAP}}$ gate while still allowing independent initialization and measurement when the exchange is low.

The first two-qubit gate demonstrated in silicon was a CZ gate performed in isotopically pure ^{28}Si metal–oxide–semiconductor quantum dots[6], using magnetic control with an on-chip electron spin resonance (ESR) antenna in 480 ns. By combining the CZ operation with a single-qubit operation, a CNOT gate was realized in about 1.2 μs, albeit without independent single-shot spin readout. Given that ESR requires a strong oscillating magnetic field, Zajac et al.[7] pursued the integration of a micromagnet in ^{nat}SiGe (^{nat}Si, natural silicon) quantum dots to use faster electrical pulses rather than magnetic control[15]. As a consequence, they demonstrated a CNOT gate via a single microwave pulse (CROT) in about 200 ns (nearly an order of magnitude faster than the CZ gate) with a 75% Bell state fidelity. In 2018, Watson et al.[8], also demonstrated a CNOT gate in ^{nat}SiGe quantum dots using electrically driven spin resonance (EDSR) within 280 ns with an average Bell state fidelity of 85%–89% (after removing the readout errors on the qubits). Recently, Huang et al.[10] used a CROT gate and ESR to generate Bell states with an average fidelity of 85% within 1.4 μs, after removing readout errors.

It is well known that the main source of decoherence when using the exchange interaction is charge noise along the detuning axis between two qubits[16]. Nevertheless, Nowack et al.[17] were able to demonstrate independent spin readout combined with sufficient control over the exchange interaction to perform a $\sqrt{\text{SWAP}}$ operation in GaAs gate-defined quantum dots. In this paper we present the first demonstration of a $\sqrt{\text{SWAP}}$ two-qubit gate in silicon between independently measured electron spins bound to phosphorus donors in natural silicon. In the long term, by combining the small size and excellent coherence properties of atom-scale qubits in silicon, we aim to utilize the hallmark long coherence times that are normally associated with ion trap qubits together with the scalability of the silicon material system to realize a large-scale quantum processor.

The two-qubit device (Fig. 1a) was fabricated using scanning tunnelling microscopy (STM) hydrogen lithography for precision placement of the donors and readout structures[18,19] (see Methods). The qubits are weakly tunnel-coupled to a radiofrequency single-electron transistor (RF-SET) that acts as a charge sensor and electron reservoir to load the electrons onto the donor quantum dots (left (L) and right (R); see Fig. 1b). Three electrical gates (left, middle and right) are used to control the electrochemical potentials of the quantum dots, whereas the SET gate is predominantly used to control the electrochemical potential of the RF-SET. The donor-based quantum dots consist of separate clusters of 2P and 3P atoms for L and R, respectively, determined by their charging energies (see Methods). The reasons for atomically engineering asymmetry in the number of donors in each quantum dot is to extend spin relaxation times[20], mitigate spatial exchange oscillations[21] and increase the tunability of the exchange interaction[22]. The device was measured in a dilution refrigerator with a base temperature of 50 mK (see Methods) and a Zeeman energy of $E_Z/h = \gamma_e B/2\pi \approx 70$ GHz (magnetic field $B = 2.5$ T; h, Planck's constant; γ_e, electron gyromagnetic ratio).

We use the (1, 3) electron charge states (where (n_L, n_R) are the electron numbers of L and R, respectively) to operate our qubits, where the 3P donor hyperfine energy is reduced by electronic shielding from the inner electrons[23]. The $\sqrt{\text{SWAP}}$ gate is performed at the (1, 3) ↔ (2, 2) charge transition shown in Fig. 1c, which is equivalent to the

[1]Centre of Excellence for Quantum Computation and Communication Technology, School of Physics, University of New South Wales, Sydney, New South Wales, Australia. [2]These authors contributed equally: Y. He, S. K. Gorman. *e-mail: michelle.simmons@unsw.edu.au

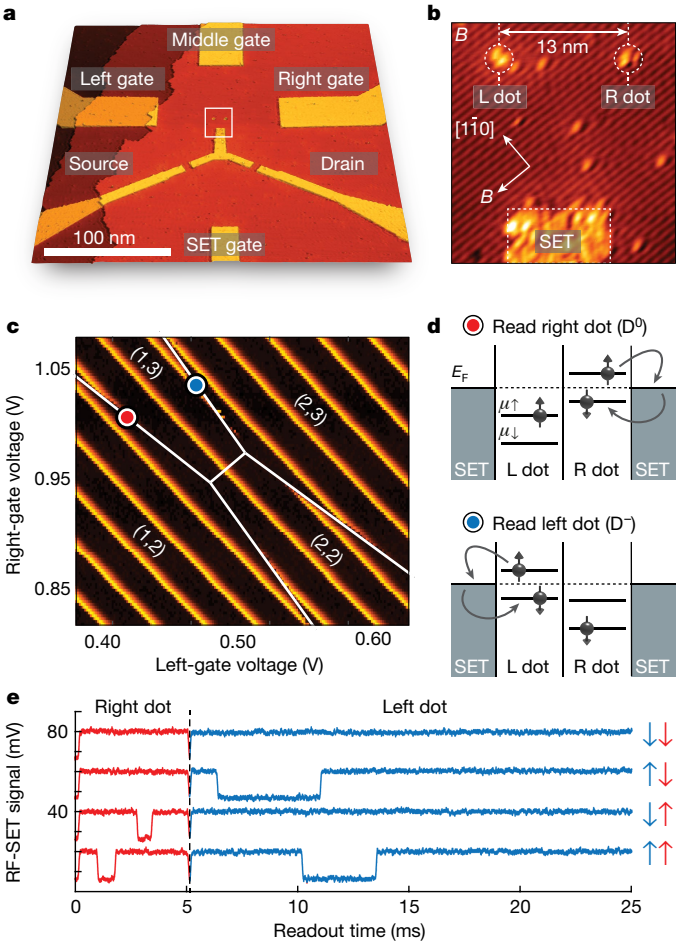


Fig. 1 | High-fidelity, independent single-shot spin readout of two donor qubits. a, STM micrograph of the two-qubit device. The lighter regions show the open lithographic hydrogen mask. The device consists of four gates: left, middle, right and the SET gate used to control the electrochemical potentials of the qubits and the RF-SET. **b**, Close-up STM micrograph of the RF-SET and the two donor dots that define the qubits L (left) and R (right). The donor quantum dots (L, 2P; R, 3P) are separated by 13.0 ± 0.5 nm. **c**, Reflected amplitude of the RF-SET as a function of the left- and right-gate voltages with electron numbers (n_L, n_R). The white lines indicate charge transitions between individual qubits and the RF-SET. The blue and red dots mark the position where spin readout is performed on the left and right qubit, respectively. **d**, Electrochemical potentials $\mu_\uparrow$ and $\mu_\downarrow$ of the two qubits (with spin $|\uparrow\rangle$ and $|\downarrow\rangle$, respectively), showing how the electron spin states are measured using an energy-selective protocol. For R we use the standard D^0-style readout and for L we use a D$^-$-style readout[24], which involves conditionally loading an electron onto the quantum dot. **e**, Reflected RF-SET amplitude measured as a function of time, used to distinguish between the four possible qubit states. The rising (falling) edge of events in the readout trace corresponds to $|\uparrow\rangle$ ($|\downarrow\rangle$) electron tunnelling to (from) the RF-SET.

(1, 1) ↔ (2, 0) transition with an inactive singlet state on R. At this higher charge transition we have a faster tunnel rate from the SET reservoir to the quantum dots so we can minimize the sequential spin readout time, enabling the collection of more statistics for mapping out the exchange oscillations. Single-shot electron spin measurements are performed using energy-selective readout[24] on each quantum dot. Spin readout relies on aligning the SET Fermi level between spin up ($|\uparrow\rangle$) and spin down ($|\downarrow\rangle$) energy; see Fig. 1d. If the electron is spin-up, then it will tunnel to the reservoir and cause a dip in the RF-SET signal. A new electron in the spin-down state will then tunnel onto the quantum dot and the RF-SET response will return to the idle value (D^0-style readout). If the electron is initially spin-down, then no tunnel events will occur and the response of the RF-SET sensor will not dip. A similar process is used on L, where, instead of unloading an electron, we conditionally load an electron onto the quantum dot, which produces the same response in the SET (D$^-$-style readout). In Fig. 1e we show four readout traces, which are assigned to $|\downarrow\downarrow\rangle$, $|\uparrow\downarrow\rangle$, $|\downarrow\uparrow\rangle$ and $|\uparrow\uparrow\rangle$. The measured spin readout fidelity is 96.5% ± 1.1% for L and 91.8% ± 1.5% for R, with a total sequential measurement fidelity of 93.9% ± 1.3% (see Supplementary Information section I). The relaxation time of $|\uparrow\rangle$, T_1, was measured to be 0.48 ± 0.08 s for L and 0.11 ± 0.03 s for R, consistent with previous theoretical and experimental results[25]. We also measure a tunnel coupling of $t_c/h = 4.3 \pm 0.4$ GHz, which we define as half of the energy gap of the $S_{(1,3)}$ and $S_{(2,2)}$ anticrossing, by fitting to the so-called 'spin-funnel' shown in Fig. 2a–c. This t_c has been engineered to be considerably larger than that of previously measured devices[19] by combining smaller inter-dot separation and the presence of the additional electron on the 3P quantum dot (see Supplementary Information section II).

To demonstrate coherent control over the exchange interaction required to perform a $\sqrt{\mathrm{SWAP}}$ gate we first show classical correlations between the two qubits[17,19]. We use the pulse sequence shown in Fig. 2d, e to investigate the spin correlations between the antiparallel spin states, $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$, as a function of detuning. We note that the electrons are initialized into a mixture of $|\uparrow\downarrow\rangle$ and $|\downarrow\downarrow\rangle$ in the (1, 3) charge region because it is not possible to deterministically load $|\uparrow\rangle$ electrons. We then apply a voltage pulse ε along the detuning axis to control the strength of J. To investigate our ability to switch J on and off, we vary our detuning pulse amplitude towards the (1, 3) ↔ (2, 2) anticrossing while waiting for 5 ms (much longer than the electron spin dephasing time, T_2^*) and measure the resulting two-spin probabilities P; see Fig. 2f. At very negative detuning, deep in the (1, 3) charge region, the electron spins are separate and remain in the initial state, $\rho_i = (|\uparrow\downarrow\rangle\langle\uparrow\downarrow| + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|)/2$ because $P_{\uparrow\downarrow} \approx P_{\downarrow\downarrow} \approx 50\%$. As the detuning pulse approaches $\varepsilon = 0$ mV, we observe $P_{\uparrow\downarrow} \to 25\%$ and $P_{\downarrow\uparrow} \to 25\%$ as expected, which is an indication of the onset of the exchange interaction because J causes rotation between these two states[17]. By contrast, $P_{\downarrow\downarrow}$ and $P_{\uparrow\uparrow}$ remain in their initial population because these states are unaffected by the exchange interaction.

We now investigate the temporal control over the exchange by fixing the detuning position and varying the duration of the exchange pulse to determine the decay of the oscillations between $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$. Here we use the pulse scheme with the corresponding energy level diagram in Fig. 3a–c. The large amplitude of about 40 mV, combined with the fast rise time of about 100 ps, creates a near perfect non-adiabatic pulse. In Fig. 3d we show normalized oscillations in $P_{\uparrow\downarrow}$ at various detuning positions with frequency $\hbar\Omega = \sqrt{\Delta E_Z^2 + J(\varepsilon, t_c)^2}$, which corresponds to the difference in energy between the T_0 state and the low-lying singlet state in the range 130–580 MHz. We can immediately see the effect of charge noise on the decay of the oscillations. As J becomes larger, the oscillations decay more rapidly owing to the larger effective noise in J because[16] $T_2^{\mathrm{SWAP}} \propto \mathrm{d}J/\mathrm{d}\varepsilon$. By fitting the oscillations to a quasi-static charge noise model we estimate the detuning noise in our device, which is determined to be $\sigma_\varepsilon \approx 16-100$ μeV at an electron temperature of 330 mK depending on the assumed curvature of J near $\varepsilon = 0$ mV (see Supplementary Information section III). Because the charge noise is known to scale with the electron temperature[16], future work will focus on reducing the electron temperature of our devices.

Finally, we demonstrate the $\sqrt{\mathrm{SWAP}}$ gate, where we use the pulse scheme shown in Fig. 4a, b to generate the input states $\rho_{\downarrow\downarrow} = |\downarrow\downarrow\rangle\langle\downarrow\downarrow|$, $\rho_{\uparrow\downarrow} = (|\uparrow\downarrow\rangle\langle\uparrow\downarrow| + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|)/2$, $\rho_{\downarrow\uparrow} = ((|\downarrow\uparrow\rangle\langle\downarrow\uparrow| + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|))/2$ and $\rho_{\mathrm{RR}} = (|\uparrow\uparrow\rangle\langle\uparrow\uparrow| + |\uparrow\downarrow\rangle\langle\uparrow\downarrow| + |\downarrow\uparrow\rangle\langle\downarrow\uparrow| + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|)/4$ (see Supplementary Information section IV). For the $\sqrt{\mathrm{SWAP}}$ gate we use $\Omega/2\pi \approx 300$ MHz because for this value of exchange we achieve an optimal trade-off between long coherence time (small $\mathrm{d}J/\mathrm{d}\varepsilon$) and high visibility oscillations (large $J > \Delta E_Z$). For initial states $\rho_{\uparrow\downarrow}$ and $\rho_{\downarrow\uparrow}$ we see oscillations that are π out of phase for $P_{\uparrow\downarrow}$ and $P_{\downarrow\uparrow}$ owing to the exchange interaction. After about 0.8 ns, $P_{\uparrow\downarrow} \approx P_{\downarrow\uparrow}$, demonstrating a $\sqrt{\mathrm{SWAP}}$ gate; at about 1.6 ns, $P_{\uparrow\downarrow}$ and $P_{\downarrow\uparrow}$ have switched values, showing that a SWAP gate has been achieved. As expected, $|\downarrow\downarrow\rangle$ and $|\uparrow\uparrow\rangle$

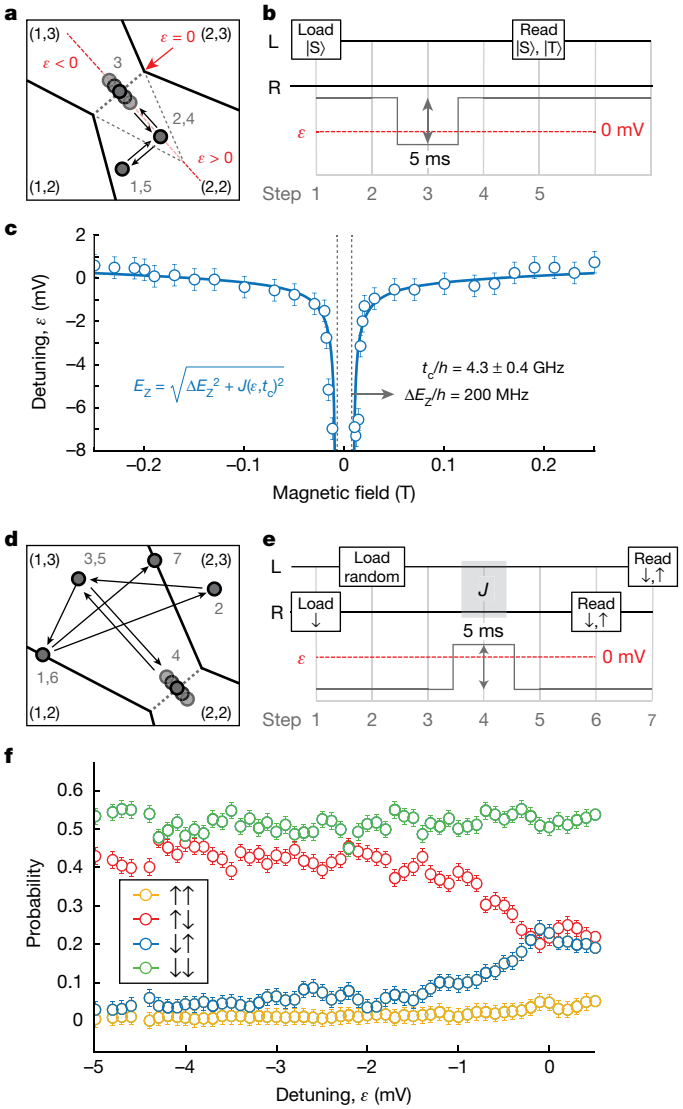


Fig. 2 | Electrostatic control over the electron-exchange interaction. **a**, **b**, Spin funnel measurement protocol. The voltage pulse (rise time of about 10 μs) along the detuning axis (dashed red line) is followed by a readout phase that allows us to distinguish between the singlet and triplet states[18]. **c**, Spin funnel measurement. The anticrossing point between $S_{(1,3)}$ and T_- is mapped as a function of detuning ε and magnetic field. By fitting the shape of the spin funnel to an equation of the form $E_Z = \sqrt{\Delta E_Z^2 + J(\varepsilon, t_c)^2}$ we obtain the difference in Zeeman energy between two electrons, $\Delta E_Z/h = 0.2 \pm 0.1$ GHz, and the tunnel coupling, $t_c/h = 4.3 \pm 0.4$ GHz. **d**, **e**, Two-spin correlation measurement protocol. A random spin is loaded onto L while R is deterministically prepared as spin-down. This preparation step is then followed by a 5-ms exchange pulse (rise time of about 10 μs) at a given detuning voltage ε, which is varied in this experiment. Finally, the spins on both quantum dots are measured. **f**, Measured two-spin probabilities of the left and right quantum dots. As the detuning pulse approaches $\varepsilon = 0$, $P_{|\uparrow}$ and $P_{|\downarrow}$ become equal owing to the increasing exchange interaction between the $S_{(1,3)}$ and T_0 states. The $|\downarrow\downarrow\rangle$ and $|\uparrow\uparrow\rangle$ probabilities remain constant, as the T_+ and T_- states are not affected by the exchange interaction because Zeeman splitting is much larger than the exchange energy for all $\varepsilon \lesssim 0$ mV ($B = 2.5$ T). The error bars represent uncertainty of one standard deviation in the measured values.

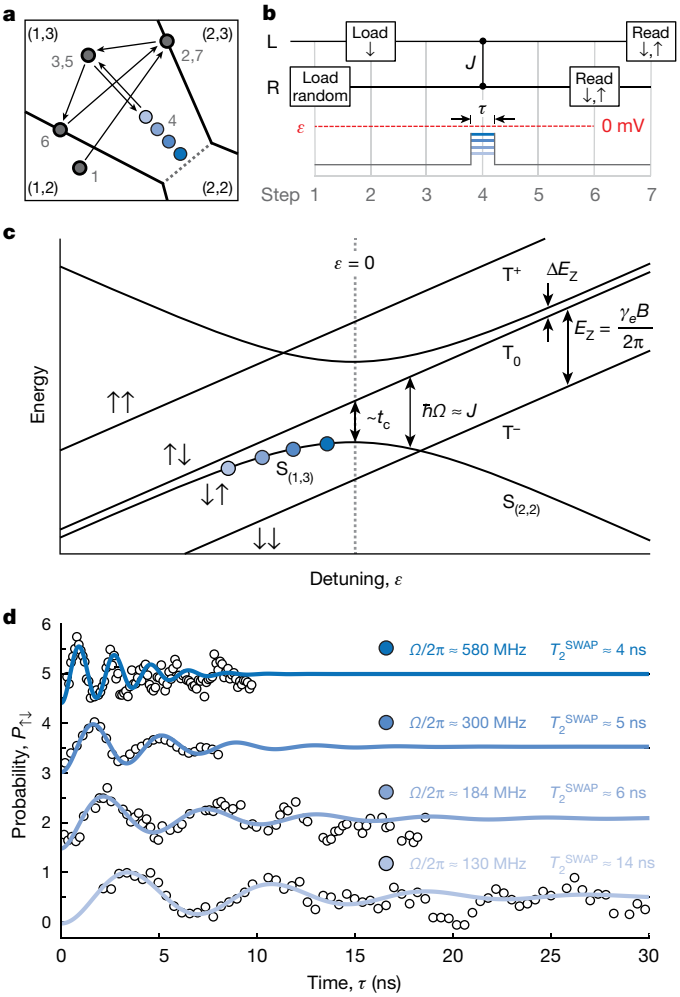


Fig. 3 | Exchange-driven coherent spin–spin oscillations. **a**, **b**, Pulse sequence employed to observe the coherent exchange oscillations. The sequence is similar to that shown in Fig. 2d, e, but the detuning position is fixed at various values (shown by the circles in **c**) and the duration of the exchange pulse is varied up to 30 ns with a rise time of about 100 ps. **c**, Energy diagram showing the evolution of singlet and triplet levels as a function of detuning. **d**, Measured probabilities of the $|\uparrow\downarrow\rangle$ spin state; error bars (1 s.d.) are smaller than the marker size. The four datasets were taken at different detuning voltages, marked with blue circles in **c** (offset by 1.5 for clarity). The T_2^{SWAP} times were determined from fits of the data (solid lines) to the results of the charge noise model (see Supplementary Information section III). We note the deviations from theory around 14 ns ($\Omega/2\pi \approx 184$ MHz) and 20 ns ($\Omega/2\pi \approx 130$ MHz), which can be explained by charge noise coupling to the SET charge sensor shifting the optimal readout position.

show no oscillations regardless of the input state. The completely random initial state, ρ_{RR}, shows no oscillations because the equally probable $|\uparrow\downarrow\rangle$ and $|\downarrow\uparrow\rangle$ oscillations cancel each other out. As a result, we observe near-flat outputs with evenly distributed populations. The deviations from the expected measurement probabilities for the different input states are attributed to initialization and readout errors.

On the basis of the four different initial-state measurements, we calculate the truth table for the $\sqrt{\text{SWAP}}$, SWAP and SWAP^2 (identity) gates after removing initialization and readout errors (see Fig. 4g and Supplementary Information section V). The truth table gives the input and output of the gate for the four different spin basis states, $\{|\downarrow\downarrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\uparrow\uparrow\rangle\}$. To quantitatively analyse the gate performance we calculate the logical basis fidelity, F_{zz} (the gate fidelity in the z basis of both qubits), using the experimental and theoretical truth table matrices[26] $E = (E_{ij})$ and $T = (T_{ij})$. The logical basis fidelities for the $\sqrt{\text{SWAP}}$, SWAP and SWAP^2 gates are $F_{zz,\sqrt{S}} = 90\% \pm 3\%$, $F_{zz,S} = 79\% \pm 3\%$ and $F_{zz,S^2} = 69\% \pm 2\%$, respectively. These fidelities provide an upper bound on the overall two-qubit gate fidelity[26], $F \leq F_{zz}$, which we estimate based on theoretical calculations to be $F_{\sqrt{S}} \approx 86.7\% \pm 0.2\%$. Using theoretical process tomography calculations we estimate that $\sigma_\varepsilon < 10$ μeV should be sufficient to achieve a $\sqrt{\text{SWAP}}$ fidelity greater than 99%, which we believe is achievable

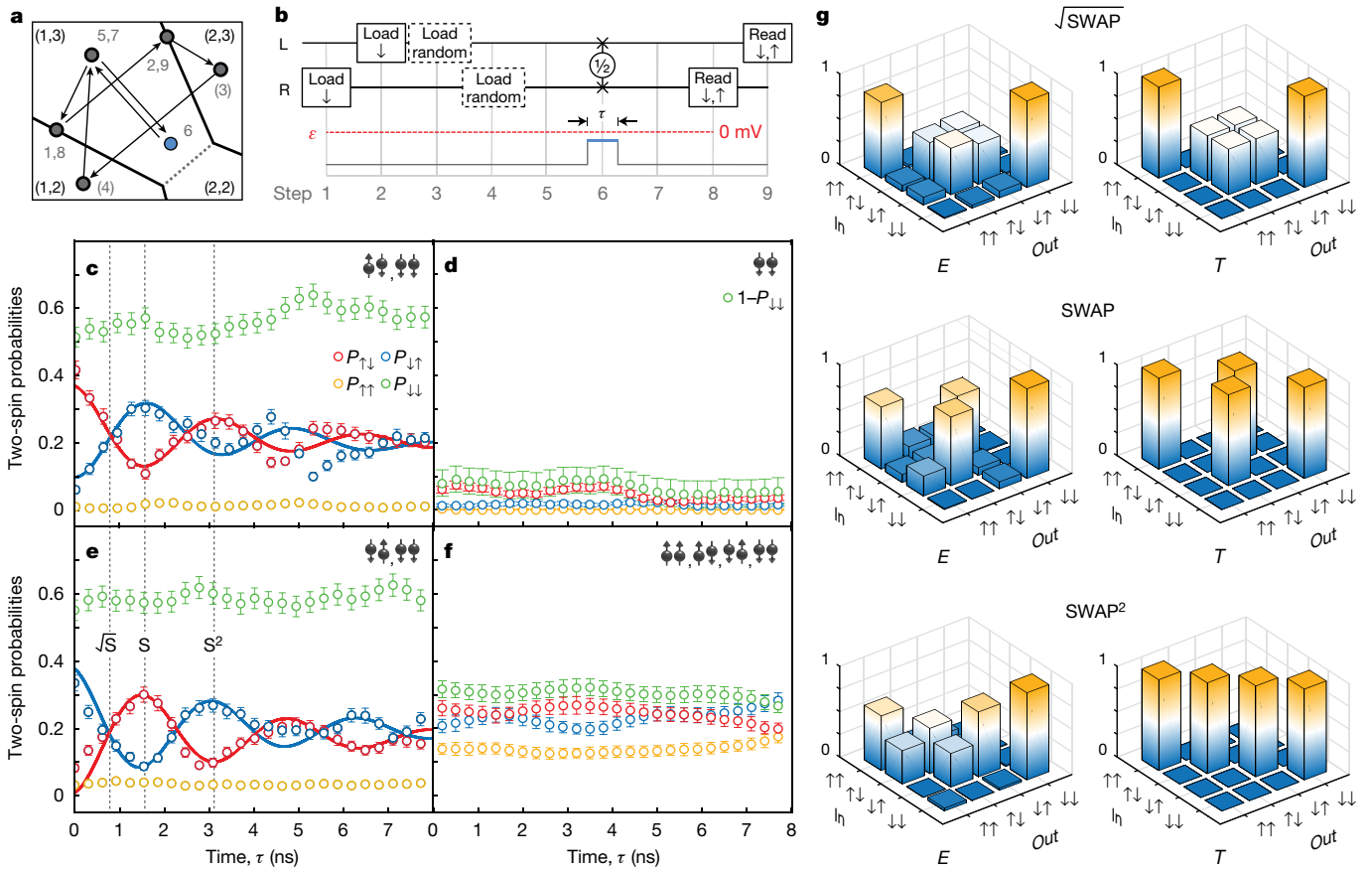


Fig. 4 | Two-qubit SWAP gate with truth table. a, **b**, Pulse sequence for performing a SWAP gate using a voltage pulse with a rise time of about 100 ps. The different input states are initialized by varying the loading position on each qubit (see Supplementary Information section IV). **c**–**f**, Measured two-spin probabilities in the basis $\{|\downarrow\downarrow\rangle, |\uparrow\downarrow\rangle, |\downarrow\uparrow\rangle, |\uparrow\uparrow\rangle\}$ with initial states $\rho_{\downarrow\downarrow} = |\downarrow\downarrow\rangle\langle\downarrow\downarrow|$, $\rho_{\uparrow\downarrow} = (|\uparrow\downarrow\rangle\langle\uparrow\downarrow| + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|)/2$, $\rho_{\downarrow\uparrow} = (|\downarrow\uparrow\rangle\langle\downarrow\uparrow| + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|)/2$ and $\rho_{RR} = (|\uparrow\uparrow\rangle\langle\uparrow\uparrow| + |\uparrow\downarrow\rangle\langle\uparrow\downarrow| + |\downarrow\uparrow\rangle\langle\downarrow\uparrow| + |\downarrow\downarrow\rangle\langle\downarrow\downarrow|)/4$, respectively. All data are measured with an exchange coupling of $\Omega/2\pi \approx 300$ MHz (second trace from the top in Fig. 3d). In **c** and **e**, the solid blue and red lines are the fits to $P_{\uparrow\downarrow}$ and $P_{\downarrow\uparrow}$. **g**, Results (E) of the truth tables of the $\sqrt{\text{SWAP}}$, SWAP and SWAP^2 gates compared with the corresponding ideal cases (T). The $\sqrt{\text{SWAP}}$ data are extracted from the two-spin probabilities at the $\pi/2$ exchange oscillation ($t = 0.77$ ns), as indicated by the dotted line labelled $\sqrt{S}$ in **c**, **e**. The SWAP gate (S in **c**, **e**) is completed after a π oscillation at $t = 1.54$ ns. Finally, the full 2π oscillation (S^2 in **c**, **e**) occurs at $t = 3.08$ ns, which results in an identity operation. The error bars represent uncertainty of one standard deviation in the measured values.

through optimizing device fabrication processes (see Supplementary Information section III). In the future, by using single-qubit rotations in the rotated basis we will be able to extract the overall two-qubit gate fidelity. This will require transitioning to devices made in isotopically purified[27 28]Si, in which high-fidelity single-qubit gates have been demonstrated[2].

The results presented in this paper demonstrate the first two-qubit gate for coupled donor atom qubits in silicon. The $\sqrt{\text{SWAP}}$ gate fidelity in the z basis, $F_{zz,\sqrt{S}} = 90\% \pm 3\%$ is ultimately limited by charge noise along the detuning axis of the qubits, which controls the strength of the exchange interaction[28]. Several methods have been proposed to reduce charge noise, such as using symmetric gate operations[29], applying composite pulse sequences[30] and designing a device with separated RF-SET and electron reservoir to reduce back-action of the charge sensor[31]. These possibilities, combined with the recently demonstrated low charge noise in buried planar devices compared with other two-dimensional materials[32] and methodologies to improve device crystallinity[33], bode well for donor-based $\sqrt{\text{SWAP}}$ gates. Future experiments will focus on measuring the Bell states obtained using the $\sqrt{\text{SWAP}}$ gate to demonstrate entanglement between two electrons using isotopically purified ^{28}Si.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1381-2.

Received: 22 October 2018; Accepted: 28 May 2019;
Published online 17 July 2019.

1. Kane, B. E. A silicon-based nuclear spin quantum computer. *Nature* **393**, 133–137 (1998).
2. Muhonen, J. T. et al. Storing quantum information for 30 seconds in a nanoelectronic device. *Nat. Nanotechnol.* **9**, 986–991 (2014).
3. Muhonen, J. T. et al. Quantifying the quantum gate fidelity of single-atom spin qubits in silicon by randomized benchmarking. *J. Phys. Condens. Matter* **27**, 154205 (2015).
4. Hill, C. D. et al. Global control and fast solid-state donor electron spin quantum computing. *Phys. Rev. B* **72**, 045350 (2005).
5. Loss, D. & DiVincenzo, D. P. Quantum computation with quantum dots. *Phys. Rev. A* **57**, 120–126 (1998).
6. Veldhorst, M. et al. A two-qubit logic gate in silicon. *Nature* **526**, 410–414 (2015).
7. Zajac, D. M. et al. Resonantly driven CNOT gate for electron spins. *Science* **359**, 439–442 (2018).
8. Watson, T. F. et al. A programmable two-qubit quantum processor in silicon. *Nature* **555**, 633–637 (2018).
9. Brunner, R. et al. Two-qubit gate of combined single-spin rotation and interdot spin exchange in a double quantum dot. *Phys. Rev. Lett.* **107**, 146801 (2011).
10. Huang, W. et al. Fidelity benchmarks for two-qubit gates in silicon. *Nature* **569**, 532–536 (2019).
11. Meunier, T., Calado, V. E. & Vandersypen, L. M. K. Efficient controlled-phase gate for single-spin qubits in quantum dots. *Phys. Rev. B* **83**, 121403 (2011).
12. Kalra, R., Laucht, A., Hill, C. D. & Morello, A. Robust two-qubit gates for donors in silicon controlled by hyperfine interactions. *Phys. Rev. X* **4**, 021044 (2014).
13. Hile, S. J. et al. Radio frequency reflectometry and charge sensing of a precision placed donor in silicon. *Appl. Phys. Lett.* **107**, 093504 (2015).
14. Weber, B. et al. Spin blockade and exchange in Coulomb-confined silicon double quantum dots. *Nat. Nanotechnol.* **9**, 430–435 (2014).

15. Yoneda, J. et al. A quantum-dot spin qubit with coherence limited by charge noise and fidelity higher than 99.9%. *Nat. Nanotechnol.* **13**, 102–106 (2018).
16. Dial, O. E. et al. Charge noise spectroscopy using coherent exchange oscillations in a singlet-triplet qubit. *Phys. Rev. Lett.* **110**, 146804 (2013).
17. Nowack, K. C. et al. Single-shot correlations and two-qubit gate of solid-state spins. *Science* **333**, 1269–1272 (2011).
18. Broome, M. A. et al. High-fidelity single-shot singlet-triplet readout of precision-placed donors in silicon. *Phys. Rev. Lett.* **119**, 046802 (2017).
19. Broome, M. A. et al. Two-electron spin correlations in precision placed donors in silicon. *Nat. Commun.* **9**, 980 (2018).
20. Hsueh, Y.-L. et al. Spin-lattice relaxation times of single donors and donor clusters in silicon. *Phys. Rev. Lett.* **113**, 246406 (2014).
21. Koiller, B., Hu, X. & Das Sarma, S. Exchange in silicon-based quantum computer architecture. *Phys. Rev. Lett.* **88**, 027903 (2001).
22. Wang, Y. et al. Highly tunable exchange in donor qubits in silicon. *npj Quantum Inf.* **2**, 16008 (2016).
23. Wang, Y., Chen, C.-Y., Klimeck, G., Simmons, M. Y. & Rahman, R. Characterizing Si:P quantum dot qubits with spin resonance techniques. *Sci. Rep.* **6**, 31830 (2016); corrigendum **6**, 38120 (2016).
24. Watson, T. F., Weber, B., House, M. G., Büch, H. & Simmons, M. Y. High-fidelity rapid initialization and read-out of an electron spin via the single donor D^- charge state. *Phys. Rev. Lett.* **115**, 166806 (2015).
25. Watson, T. F. et al. Atomically engineered electron spin lifetimes of 30 s in silicon. *Sci. Adv.* **3**, e1602811 (2017).
26. Politi, A., Cryan, M. J., Rarity, J. G., Yu, S. & O'Brien, J. L. Silica-on-silicon waveguide quantum circuits. *Science* **320**, 646–649 (2008).
27. Abrosimov, N. V. et al. A new generation of 99.999% enriched 28Si single crystals for the determination of Avogadro's constant. *Metrologia* **54**, 599–609 (2017).
28. Throckmorton, R. E., Barnes, E. & Das Sarma, S. Environmental noise effects on entanglement fidelity of exchange-coupled semiconductor spin qubits. *Phys. Rev. B* **95**, 085405 (2017).
29. Martins, F. et al. Noise suppression using symmetric exchange gates in spin qubits. *Phys. Rev. Lett.* **116**, 116801 (2016).
30. Wang, X. et al. Composite pulses for robust universal control of singlet–triplet qubits. *Nat. Commun.* **3**, 997 (2012).
31. Horibe, K., Kodera, T. & Oda, S. Back-action-induced excitation of electrons in a silicon quantum dot with a single-electron transistor charge sensor. *Appl. Phys. Lett.* **106**, 053119 (2015).
32. Shamim, S., Weber, B., Thompson, D. W., Simmons, M. Y. & Ghosh, A. Ultra low-noise atomic-scale structures for quantum circuitry in silicon. *Nano Lett.* **16**, 5779–5784 (2016).
33. Keizer, J. G., Koelling, S., Koenraad, P. M. & Simmons, M. Y. Suppressing segregation in highly phosphorus doped silicon monolayers. *ACS Nano* **9**, 12537–12541 (2015).

Acknowledgements The research was supported by the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (project number CE170100012), the US Army Research Office under contract number W911NF-17-1-0202 and Silicon Quantum Computing Pty Ltd. M.Y.S. acknowledges an Australian Research Council Laureate Fellowship. This work was performed in part at the NSW node of the Australian National Fabrication Facility.

Reviewer information *Nature* thanks Benjamin D'Anjou and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Y.H., S.K.G. and L.K. fabricated the device. Y.H., S.K.G. and D.K. performed the measurements. Y.H., S.K.G., D.K, L.K. and J.G.K. analysed the data. The manuscript was written by Y.H., S.K.G. and M.Y.S. with input from all other authors. M.Y.S. conceived and supervised the project.

Competing interests M.Y.S. is a director of the company Silicon Quantum Computing Pty Ltd.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1381-2.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1381-2.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to M.Y.S.

METHODS

Device fabrication and choice of quantum dot size. The device is fabricated on a p-type natural Si substrate (1–10 Ω cm). The substrate is subjected to a series of high-temperature annealing processes up to 1,100 °C followed by a controlled cool-down to 330 °C, at which point the surface is terminated with mono-atomic hydrogen via thermal cracking. The result is a fully terminated H:Si (2 × 1) reconstructed surface from which the hydrogen can be selectively removed with the STM tip. Using the STM tip a lithographic mask representing the device and donor qubits is created on the Si surface. Subsequent adsorption and incorporation (at 350 °C) of gaseous PH_3 precursor metallizes the exposed area with ~1/4 monolayer of phosphorus. Then, 45 nm of natural Si is grown on the device using molecular beam epitaxy.

The charging energy of the quantum dots, E_c^D, can be used to determine the number of donors in each qubit[14,19]. In Extended Data Fig. 1 we show the definitions of the parameters used for this calculation. The first is the RF-SET charging energy, E_c^S, which is measured from the Coulomb diamonds in Extended Data Fig. 1a. The mutual charging energy between the RF-SET and the donor qubit gives a voltage shift on the RF-SET transition in the gate map that is defined as δV_g^S; see Extended Data Fig. 1b. Its counterpart, the donor transition shift, is defined as δV_g^D. Similarly, the voltage shift on donor transitions caused by the mutual charging energy between the donors is defined as δV_g^D. The voltage differences between the two charge transitions of the RF-SET (donor) is defined as ΔV_g^S (ΔV_g^D), shown in Extended Data Fig. 1b (Extended Data Fig. 1c). Finally, the number of charge transitions of the SET between the two donor quantum dot charging events, n^s, can be determined by counting the number of current peaks in the charge stability maps. After measuring all of these values in the charge stability map, we can calculate the donor charging energy E_c^D, which is given by

$$E_c^D = \frac{E_c^S \delta V_g^S}{\Delta V_g^S}\left(\frac{\Delta V_g^D - \delta V_g^{DD}}{\delta V_g^D} - n^S\right) \quad (1)$$

By comparing the calculated charging energies of the L dot, $E_c^D(\mathrm{L}) = 70.9 \pm 14.7$ meV, and the R dot, $E_c^D(\mathrm{R}) = 90.2 \pm 18.5$ meV, with previous results and theoretical calculations[14], we estimate that the L dot contains 2P and the R dot contains 3P. We also check the donor numbers by measuring the spin relaxation time T_1 on both quantum dots. On L we find $T_1 = 0.483 \pm 0.083$ s and on R we determine $T_1 = 0.107 \pm 0.032$ s at $B = 2.5$ T and an electron temperature of 330 mK. These values are comparable to previous results on 2P1e (one electron bound to two phosphorus donors) and consistent with theoretical predictions for the 3P3e case at comparable magnetic fields[20].

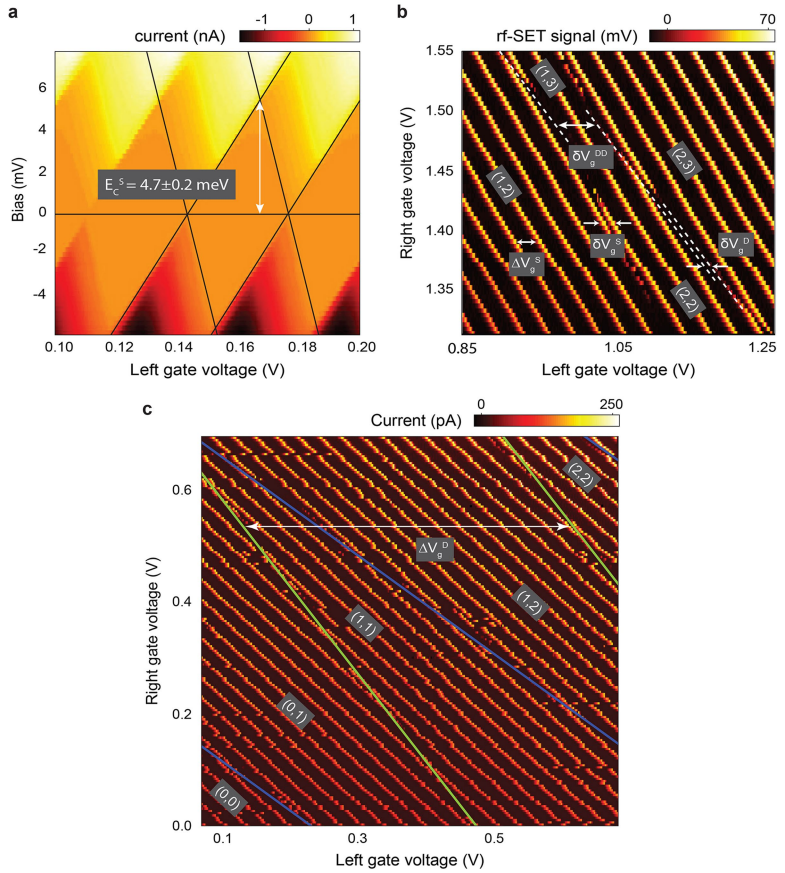
Experimental measurement set-up. The device is measured at low temperatures (50 mK) inside a dilution refrigerator; a diagram of our complete set-up is displayed in Extended Data Fig. 2. The gate electrodes (left, middle, right and SET) of the device are connected to a room-temperature voltage source (SIM 928) to control the electrostatic environment of the quantum dots, via 1:5 voltage dividers and a two-stage RC low-pass filter (~300 kHz) at the 50-mK stage. An additional voltage source (NI6363) with 1:50 voltage dividers is connected to the left and right gates to supply the voltage pulses required for initializing the qubits. Faster pulsing is required for the nanosecond detuning pulses that are needed to create coherent state rotations, so two room-temperature, synchronized waveform generators (AWG70001A) are also connected to the left and right gates through bias tees and triggered by the NI6363. The bias tees each contain a 2-kΩ resistor and a 1.2-nF capacitor, and the fast lines each have attenuators of −18 dB between the 4-K and 50-mK stages for noise reduction and thermalization.

The spin readout of the quantum dots is performed with an SET used as a charge detector comprising a quantum dot, source and drain leads. We perform readout with the SET in combination with a lumped-element LC circuit (RF-SET) in the megahertz regime using reflectometry techniques[34]. The LC circuit consists of a 1-μH inductor and the parasitic capacitance of the device (<1 pF) with a resonance frequency of 258.3 MHz, and is used to impedance-match the SET with the rest of the 50-Ω electronics. The input carrier wave (258.3 MHz) for the RF-SET readout is supplied by a room-temperature RF signal generator and is passed through a variable attenuator and d.c. block before reaching the dilution unit. The carrier wave is further attenuated between the 4-K and 50-mK stages (−40 dB) and travels through a directional coupler before being reflected by the RF-SET, amplified at both 50 mK (CMT CITLF-3) and room temperature, and then demodulated with the original carrier wave at room temperature to measure the in-phase and quadrature signals.

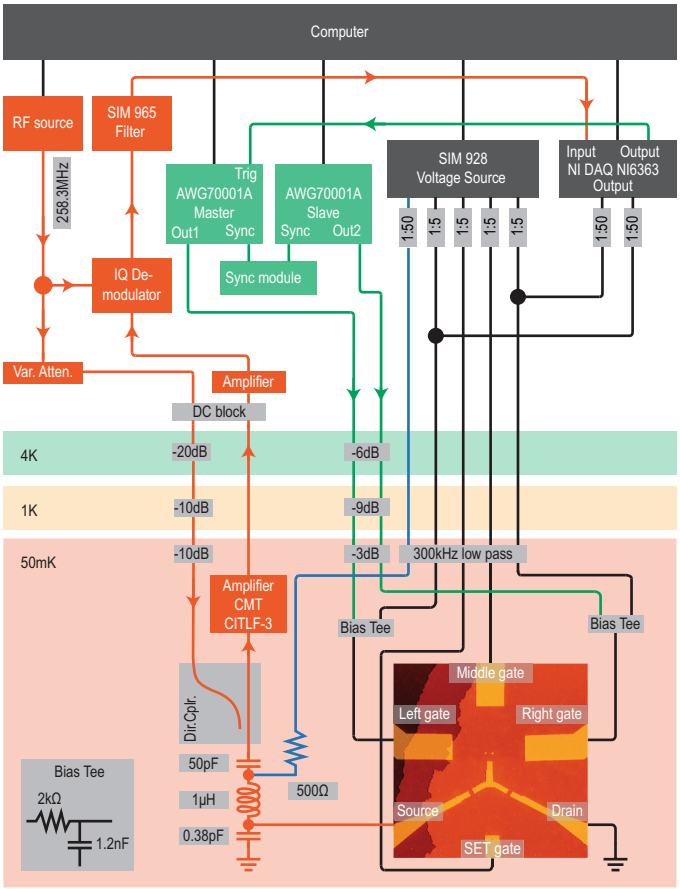
Data availability

The data pertaining to this study are available from the corresponding author upon reasonable request.

34. Gorman, S. K. et al. Tunneling statistics for analysis of spin-readout fidelity. *Phys. Rev. Appl.* **8**, 034019 (2017).



Extended Data Fig. 1 | Determination of the number of donors in a qubit from the quantum dot charging energies. a, RF-SET Coulomb diamonds. **b**, Gate–gate map around the (1, 3) ↔ (2, 2) charge transition and definitions of δV_g^{DD}, ΔV_g^{S}, δV_g^{S} and δV_g^{D}. **c**, Gate–gate map for the regime with two electrons on donor qubits and definition of ΔV_g^{D}.



Extended Data Fig. 2 | Experimental set-up for the two-qubit gate in a millikelvin dilution refrigerator. The schematic shows the electrical connections from the device to the control computer at the different temperature stages of the dilution refrigerator from the top to bottom. An STM image of the device attached to the cold finger of the refrigerator at 50 mK is shown in the lower orange pane. The RF-reflectormetry circuit attached to the device employs a variable attenuator ('Var. Atten.') to control the power coupled through a directional coupler ('Dir. Cplr.') and sent to the source contact of the SET (red lines). The blue line is for d.c. current/voltage measurements of the SET. The slow signals (black components) and fast signals (green components) are combined using bias tees at 50 mK before being sent to the left and right gate electrodes.

LETTER

https://doi.org/10.1038/s41586-019-1375-0

Giant thermal Hall conductivity in the pseudogap phase of cuprate superconductors

G. Grissonnanche[1]*, A. Legros[1,2], S. Badoux[1], E. Lefrançois[1], V. Zatko[1], M. Lizaire[1], F. Laliberté[1], A. Gourgout[1], J.-S. Zhou[3], S. Pyon[4,5], T. Takayama[4,6], H. Takagi[4,6,7,8], S. Ono[9], N. Doiron-Leyraud[1] & L. Taillefer[1,10]*

The nature of the pseudogap phase of the copper oxides ('cuprates') remains a puzzle. Although there are indications that this phase breaks various symmetries, there is no consensus on its fundamental nature[1]. Fermi-surface, transport and thermodynamic signatures of the pseudogap phase are reminiscent of a transition into a phase with antiferromagnetic order, but evidence for an associated long-range magnetic order is still lacking[2]. Here we report measurements of the thermal Hall conductivity (in the x–y plane, κ_{xy}) in the normal state of four different cuprates—$La_{1.6-x}Nd_{0.4}Sr_xCuO_4$, $La_{1.8-x}Eu_{0.2}Sr_xCuO_4$, $La_{2-x}Sr_xCuO_4$ and $Bi_2Sr_{2-x}La_xCuO_{6+\delta}$. We show that a large negative κ_{xy} signal is a property of the pseudogap phase, appearing at its critical hole doping, p^*. It is also a property of the Mott insulator at $p \approx 0$, where κ_{xy} has the largest reported magnitude of any insulator so far[3]. Because this negative κ_{xy} signal grows as the system becomes increasingly insulating electrically, it cannot be attributed to conventional mobile charge carriers. Nor is it due to magnons, because it exists in the absence of magnetic order. Our observation is reminiscent of the thermal Hall conductivity of insulators with spin-liquid states[4–6], pointing to neutral excitations with spin chirality[7] in the pseudogap phase of cuprates.

Among the different families of unconventional superconductors, magnetism and superconductivity are often closely associated[8]. A notable exception is the family of hole-doped cuprates, where superconductivity mostly coexists instead with the pseudogap phase, which is an enigmatic state of matter whose nature remains unclear[1]. The critical doping p^* (for the onset of the pseudogap phase) bears the hallmarks of an antiferromagnetic quantum critical point[2], with a sharp drop in the carrier density n from $n \approx 1 + p$ above p^* to $n \approx p$ below p^*, a resistivity linear with temperature T, and a specific heat with a $\log(1/T)$ dependence. Yet, there is no evidence for long-range magnetic order appearing at p^*. However, numerical solutions of the Hubbard model have shown that a pseudogap phase can arise from short-range antiferromagnetic correlations[9]. It has been argued that an exotic state with topological order can account for such a pseudogap and for the drop in carrier density without breaking translational symmetry[10], but the low-energy excitations of such a state have yet to be detected.

In recent years, the thermal Hall effect has emerged as a powerful probe of magnetic texture and topological excitations in insulators. On the theory side, a non-zero thermal Hall conductivity κ_{xy} was shown to arise even without long-range magnetic order, either from the spin chirality of a paramagnetic state[7] or from fractionalized (topological) excitations in a spin liquid[11]. On the experimental side, a sizeable κ_{xy} has been measured in insulators without magnetic order, such as the spin-ice system $Tb_2Ti_2O_7$ (ref. [12]) and the spin-liquid systems $RuCl_3$ (ref. [4]), volborthite[5] and Ca kapellasite[6].

In cuprates, studies of κ_{xy} have so far been limited to the superconducting state[13–15], except for the case of $YBa_2Cu_3O_y$ (YBCO) at $p = 0.11$, where κ_{xy} was measured in the field-induced normal state[16], which has charge-density-wave order[2]. See Methods for a discussion of this particular case.

Here, we investigate the thermal Hall response of the pseudogap phase via measurements of κ_{xy} in four different cuprate

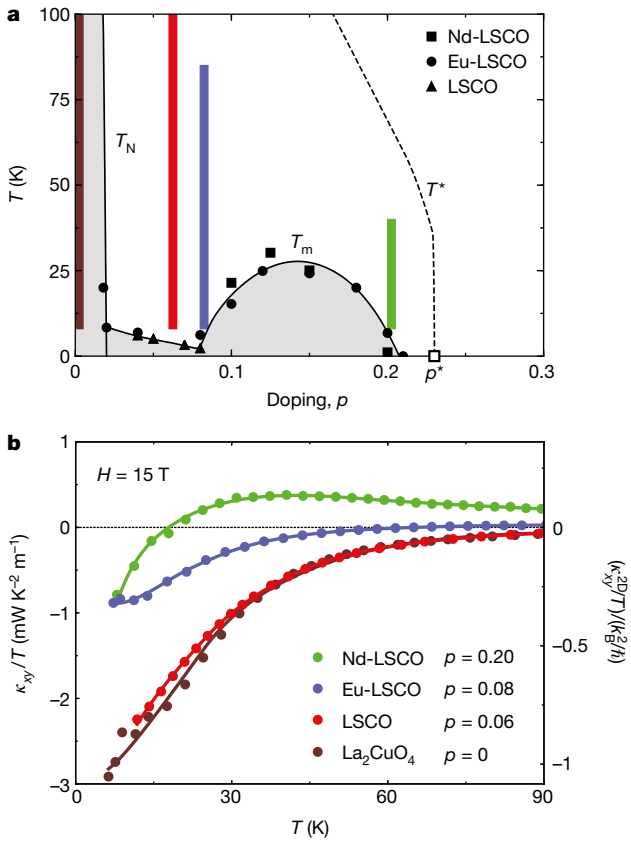


Fig. 1 | Phase diagram and thermal Hall conductivity of cuprates. **a**, Temperature–doping phase diagram of Nd-LSCO, Eu-LSCO and LSCO, showing the antiferromagnetic phase below the Néel temperature T_N and the pseudogap phase below T^* (ref. [29]), which ends at the critical doping $p^* = 0.23$ for both Nd-LSCO (ref. [17]) and Eu-LSCO (ref. [30]). For LSCO, $p^* \approx 0.18$ (ref. [29]). Short-range incommensurate spin order occurs below T_m, as measured by μSR on Nd-LSCO (squares[21]), Eu-LSCO (circles[31]) and LSCO (triangles[32]). The coloured vertical strips indicate the temperature range where the thermal Hall conductivity κ_{xy}/T at the corresponding doping decreases towards negative values at low temperature (see **b**). **b**, Thermal Hall conductivity κ_{xy}/T versus temperature in a field $H = 15$ T, for four materials and dopings as indicated, colour-coded with the vertical strips in **a**. On the right vertical axis, the magnitude of κ_{xy}/T is expressed in fundamental units of thermal conductance per plane ($k_B^2/\hbar$).

[1]Département de physique, Institut quantique, and RQMP, Université de Sherbrooke, Sherbrooke, Québec, Canada. [2]SPEC, CEA, CNRS-UMR3680, Université Paris-Saclay, Gif-sur-Yvette, France. [3]Materials Science and Engineering Program, Department of Mechanical Engineering, University of Texas at Austin, Austin, TX, USA. [4]Department of Advanced Materials Science, University of Tokyo, Kashiwa, Japan. [5]Department of Applied Physics, University of Tokyo, Tokyo, Japan. [6]Max Planck Institute for Solid State Research, Stuttgart, Germany. [7]Department of Physics, University of Tokyo, Tokyo, Japan. [8]Institute for Functional Matter and Quantum Technologies, University of Stuttgart, Stuttgart, Germany. [9]Central Research Institute of Electric Power Industry, Kanagawa, Japan. [10]Canadian Institute for Advanced Research, Toronto, Ontario, Canada. *e-mail: gael.grissonnanche@usherbrooke.ca; louis.taillefer@usherbrooke.ca

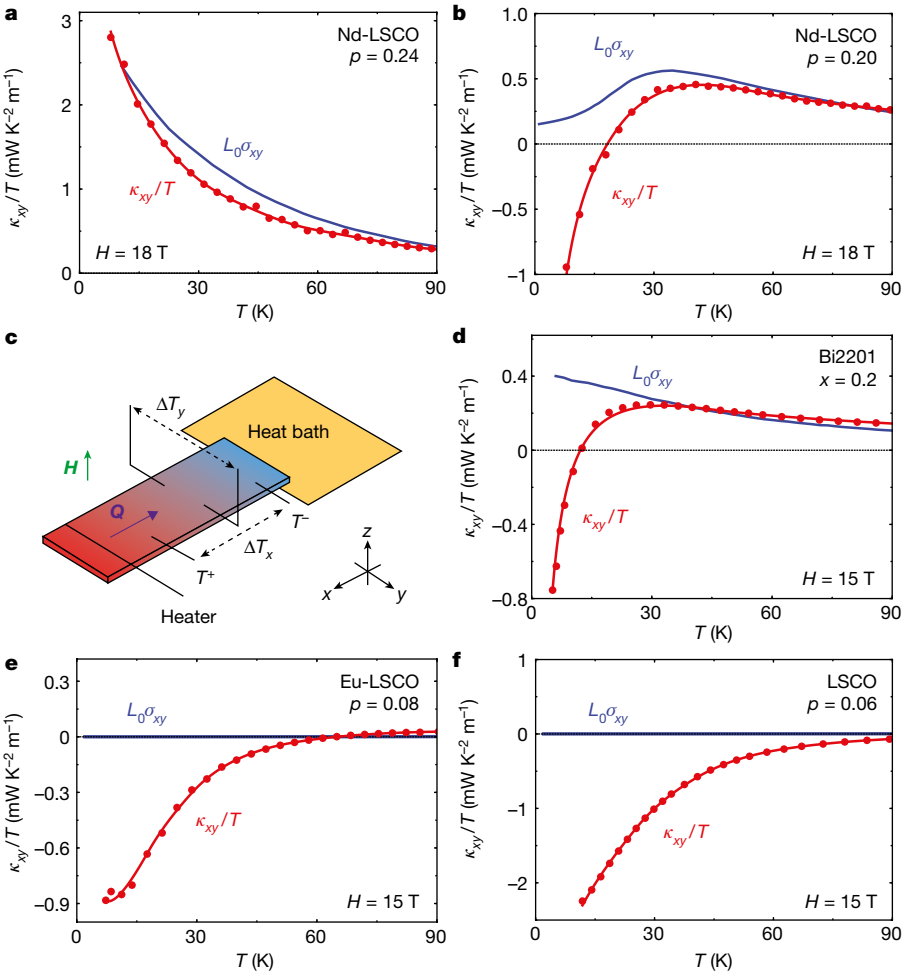


Fig. 2 | Thermal and electrical Hall conductivities of four cuprates. Data panels show thermal Hall conductivity κ_{xy}, plotted as κ_{xy}/T (red), and electrical Hall conductivity σ_{xy}, expressed as $L_0\sigma_{xy}$ (blue), where $L_0 = (\pi^2/3)(k_B/e)^2$, as a function of temperature: the material, its doping p and field H are indicated. **a**, **b**, Nd-LSCO; **c**, sketch of the thermal Hall measurement set-up (see Methods); **d**, Bi2201; **e**, Eu-LSCO; and **f**, LSCO. (For Nd-LSCO $p = 0.20$ (**b**), σ_{xy} was measured[17] at $H = 33$ T.) In Nd-LSCO at $p = 0.24$, κ_{xy}/T and $L_0\sigma_{xy}$ are both positive at all temperatures and they track each other, satisfying the Wiedemann–Franz law in the $T = 0$ limit. By contrast, for $p < p^*$ in all four materials, κ_{xy}/T falls to large and negative values at low temperature, whereas $L_0\sigma_{xy}$ remains positive.

materials—$La_{1.6-x}Nd_{0.4}Sr_xCuO_4$ (Nd-LSCO), $La_{1.8-x}Eu_{0.2}Sr_xCuO_4$ (Eu-LSCO), $La_{2-x}Sr_xCuO_4$ (LSCO) and $Bi_2Sr_{2-x}La_xCuO_{6+\delta}$ (Bi2201)—across a wide doping range, from the overdoped metal at $p = 0.24$ down to the Mott insulator at $p \approx 0$ (Fig. 1a). The κ_{xy} data reported here are all in the normal state, with superconductivity suppressed by application of a magnetic field normal to the CuO_2 planes.

In Nd-LSCO and Eu-LSCO, the critical doping[17] is at $p^* = 0.23$ (Fig. 1a). In Fig. 2a, we plot κ_{xy}/T versus T for Nd-LSCO at $p = 0.24$. We find that κ_{xy} is positive and that κ_{xy}/T increases monotonically with decreasing T, tracking closely the electrical Hall conductivity σ_{xy} measured on the same sample, satisfying the Wiedemann–Franz law as $T \to 0$, namely $\kappa_{xy}/T = L_0\sigma_{xy}$, where $L_0 = (\pi^2/3)(k_B/e)^2$ (here k_B is the Boltzmann constant and e the electron charge). The large positive value of σ_{xy} is dictated by the large Fermi surface at $p > p^*$ and its positive Hall number $n_H \approx 1 + p$ (ref. [17]). Clearly, at $p = 0.24$, κ_{xy} is entirely due to the conventional Hall effect of mobile charge carriers.

We now turn to dopings immediately below the pseudogap critical point. In Fig. 2b, we plot κ_{xy}/T versus T for Nd-LSCO at $p = 0.20$. We see a qualitatively different behaviour, with κ_{xy} becoming negative at low T. As seen in Fig. 3a, this qualitative change occurs immediately below p^*. In Eu-LSCO, the very same change occurs across p^* (Fig. 3b), from positive κ_{xy} above p^* ($p = 0.24$) to negative κ_{xy} (at low T) below p^* ($p = 0.21$), with essentially identical data to Nd-LSCO at $p = 0.24$ and $p = 0.21$. The negative κ_{xy} is therefore a property of the pseudogap phase.

We also measured κ_{xy} in Bi2201 (a cuprate with a different crystal structure to that of Nd-LSCO and Eu-LSCO), using an overdoped sample of La content $x = 0.2$, with p slightly below p^* (ref. [18]). In Fig. 2d, we see that $\kappa_{xy}(T)$ in Bi2201 displays a remarkably similar behaviour to that of Nd-LSCO and Eu-LSCO at $p < p^*$. A negative thermal Hall conductivity κ_{xy} at low temperature is therefore a generic property of the pseudogap phase, independent of material. Note that the electrical Hall conductivity σ_{xy} measured on the same samples remains positive down to $T \to 0$ (Fig. 2b, d).

We now move to much lower doping. In Fig. 1b, we see that κ_{xy}/T is still negative at low temperature in Eu-LSCO at $p = 0.08$ and in LSCO at $p = 0.06$, where in both cases σ_{xy} is positive and completely negligible (Fig. 2e, f), because the samples are almost electrically insulating at low temperature. This shows that the negative κ_{xy} signal of the pseudogap phase is not due to the conventional Hall effect of mobile charge carriers.

Magnons can be excluded as the source of this negative κ_{xy}. In the phase diagram of Fig. 1a, we delineate in grey the regions where static magnetism is detected by muon spin resonance (μSR), whether as incommensurate correlations below an onset temperature T_m or as commensurate Néel order below the Néel temperature, T_N. We see that in all three materials—Nd-LSCO at $p = 0.20$, Eu-LSCO at $p = 0.08$ and LSCO at $p = 0.06$—the negative κ_{xy} signal is present well above T_m (Fig. 1), where there is no static magnetism. Moreover, the $\kappa_{xy}(T)$ curve for La_2CuO_4 (Fig. 1b), that is, undoped LSCO with $p \approx 0$, where

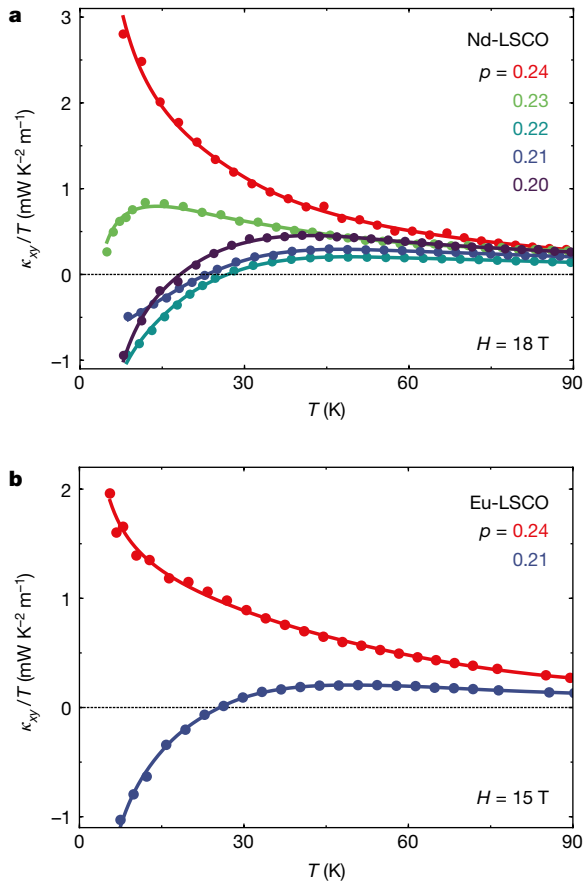


Fig. 3 | Thermal Hall conductivity across the pseudogap critical point *p.** Shown is thermal Hall conductivity κ_{xy}/T for Nd-LSCO in $H = 18$ T (**a**) and Eu-LSCO in $H = 15$ T (**b**), at dopings as indicated, on both sides of the pseudogap critical point $p^* = 0.23$. In both materials, κ_{xy} becomes negative at low temperature when $p < p^*$.

there is long-range antiferromagnetic order below approximately 300 K (Fig. 1a), is very similar to the curve for LSCO at $p = 0.06$ (Fig. 1b), where there is no magnetic order above $T \approx 5$ K (Fig. 1a). (See Methods for further discussion of magnons.) We conclude that magnetic order is not responsible for the negative κ_{xy} signal seen in cuprates at all dopings below p^*, and magnons are ruled out as the relevant excitations.

Phonons can generate a non-zero κ_{xy} signal if they are subject to scattering by spins[19,20]. Spin scattering will also show up in the longitudinal thermal conductivity κ_{xx}, which is dominated by phonons, in two ways: (1) it reduces the magnitude of κ_{xx} relative to a non-magnetic analogue material; and (2) it produces a field dependence of κ_{xx}.

In relation to (1), we note that κ_{xx} in Nd-LSCO does not decrease below p^*; on the contrary, it increases (Extended Data Fig. 3), most probably because electron–phonon scattering decreases as the charge carrier density drops. So the large negative κ_{xy} signal that appears below p^* is not accompanied by a reduction of κ_{xx} that would signal the onset of spin scattering. One could invoke a scenario where the decrease in electron–phonon scattering overcompensates the effect of the spin scattering, but the latter would still have to be small, which is hard to reconcile with the enormous κ_{xy} signal. Moreover, there is no evidence that the spin state of Nd-LSCO changes across p^*. On the contrary, static moments present at $p = 0.12$ cease to be detected (by μSR) at $p = 0.20$ (ref. [21]), so that $p = 0.20$ and $p = 0.24$ are equally non-magnetic from the μSR point of view. In other words, magnetic moments that could scatter phonons are not substantially different above and below p^*.

In relation to (2), the strength of the field (H) dependence of κ_{xx} is measured by the ratio $[\kappa_{xx}(H) - \kappa_{xx}(0)]/\kappa_{xx}(0)$. In Fig. 4a, we compare cuprates to various insulators with sizeable κ_{xy} signals. We see that the field dependence of κ_{xx} in LSCO $p = 0.06$, Eu-LSCO $p = 0.08$ and La_2CuO_4 is much smaller than in other materials, including $Ba_3CuSb_2O_9$ (ref. [20]) for example, a material where spin–phonon scattering generates the κ_{xy} signal. Although this could in part be due to a larger relevant field scale in cuprates, we are nonetheless left with little evidence of strong spin–phonon scattering in cuprates.

Given that the usual two indicators of a phonon-driven κ_{xy} are not clearly observed in our data, we conclude that phonons are unlikely to be responsible for the large negative κ_{xy} signal of cuprates that appears suddenly below p^*. (See Methods for further discussion.)

The κ_{xy} signal in the Mott insulator La_2CuO_4 is the largest seen so far in any insulator. Only multiferroic materials such as ferrimagnetic $(Fe,Zn)_2Mo_3O_8$ have comparable κ_{xy} values[3] (Fig. 4b), thanks to their exceptionally strong lattice–spin coupling—a measure of which is the strong field dependence of κ_{xx}, about 100 times larger in $(Fe,Zn)_2Mo_3O_8$ than in the cuprates (Fig. 4a).

The large negative κ_{xy} reported here for cuprates is not due to the standard Hall effect of charge carriers, it is not caused by magnons and there is no clear evidence that it comes from phonons. Its occurrence is all the more surprising given the 'no-go theorem' that should strongly limit its magnitude on a square lattice[22]. Identifying the excitations responsible for the negative κ_{xy} signal will shed new light on the nature of the pseudogap phase. It is instructive to compare cuprates with insulators that are believed to host spin-liquid states. The largest κ_{xy} signal so far in such materials was detected in $RuCl_3$ (Fig. 4b). In this 2D

Table 1 | Thermal Hall conductivity in various insulators

Material	κ_{xy} (mW K^{-1} m^{-1})	κ_{xx} (W K^{-1} m^{-1})	$\lvert\Delta\kappa_{xx}\rvert$ (W K^{-1} m^{-1})	$\lvert\Delta\kappa_{xx}/\kappa_{xx}\rvert$	T (K)	H (T)	Reference
La_2CuO_4	−38.6	12.4	~0.06	~0.005	20	15	This work
LSCO	−30.0	5.1	~0.02	~0.004	15	15	This work
Eu-LSCO	−13.2	4.5	~0.015	~0.003	15	15	This work
$Lu_2V_2O_7$	1.0	0.75	ND	ND	50	9	[28]
$Fe_2Mo_3O_8$	24	9	5	0.55	45	14	[3]
$(Fe,Zn)_2Mo_3O_8$	24	10	3.2	0.32	30	9	[3]
$Tb_2Ti_2O_7$	1.2	0.37	0.12	0.32	15.5	8	[12]
$RuCl_3$	8	15.5	0.62	0.04	20	15	[4]
$RuCl_3$	3.5	8	0.45	0.055	35	16	[23]
Ca kapellasite	1.1	0.2	ND	ND	16	15	[6]
$Ba_3CuSb_2O_9$	0.008	0.07	0.0035	0.05	5	15	[20]

Maximal value of the thermal Hall conductivity κ_{xy} (second column) in various insulators (first column), compared to our three cuprates (the first three entries, namely, La_2CuO_4, LSCO $p = 0.06$ and Eu-LSCO $p = 0.08$), measured at temperature T and field H as indicated (columns 6 and 7 respectively): the ferromagnet $Lu_2V_2O_7$ (ref. [28]); the multiferroic ferrimagnets $Fe_2Mo_3O_8$ and $(Fe_{0.875}Zn_{0.125})_2Mo_3O_8$ (ref. [3]); the spin-ice material $Tb_2Ti_2O_7$ (ref. [12]); and the spin-liquid candidates $RuCl_3$ (refs [4,23]), Ca kapellasite[6] and $Ba_3CuSb_2O_9$ (ref. [20]). We also list the thermal conductivity κ_{xx} measured at the same temperature, in zero field (third column). The change induced in κ_{xx} by the field, $\Delta\kappa_{xx} = \kappa_{xx}(H) - \kappa_{xx}(0)$, is given in absolute and relative terms (fourth and fifth column, respectively). ND, not determined.

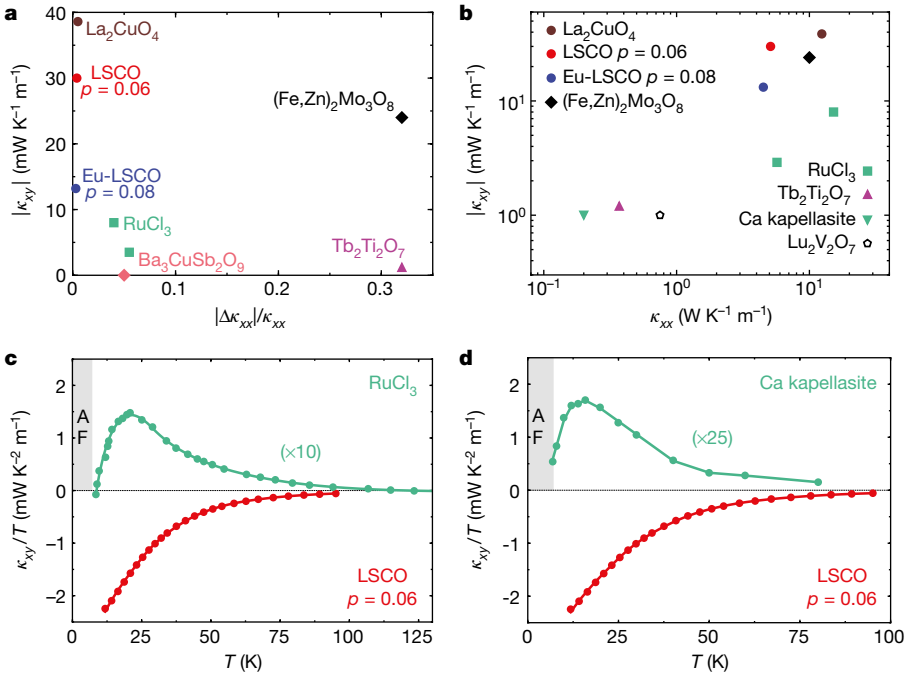


Fig. 4 | Comparison with other insulators, including spin-liquid candidates. a, **b**, Maximal absolute value of κ_{xy} in various insulators, including the multiferroic ferrimagnet $(Fe,Zn)_2Mo_3O_8$ (black diamond[3]; the previous record holder for the largest $|\kappa_{xy}|$ of any insulator) and the spin-liquid insulator $RuCl_3$ (green squares[4,23]; the previous record holder for the largest $|\kappa_{xy}|$ of any insulator without magnetic order). **a**, Maximal $|\kappa_{xy}|$ as a function of the corresponding value of $[\kappa_{xx}(H) - \kappa_{xx}(0)]/\kappa_{xx}(0)$. **b**, Maximal $|\kappa_{xy}|$ as a function of the corresponding κ_{xx} value, on a log–log plot. The values for all materials are listed in Table 1. We see that La_2CuO_4 has the largest known value of all insulators. **c**, Thermal Hall conductivity κ_{xy}/T versus temperature for LSCO at $p = 0.06$ in $H = 15$ T (red) and $RuCl_3$ in $H = 16$ T (blue, ×10; data from ref. [23]). In $RuCl_3$, the gradual growth of κ_{xy}/T on cooling below $T \approx 100$ K is attributed to Majorana fermions, the topological excitations of the Kitaev spin liquid[4,6,11]. Below $T \approx 20$ K, κ_{xy}/T drops on approaching the antiferromagnetic phase (AF; grey). **d**, Same as in **c** but for the spin-liquid insulator Ca kapellasite (green, ×25; data from ref. [6]). These comparisons point to a spin-liquid character of the pseudogap phase in cuprates.

material, spins on a honeycomb lattice are frustrated and only order (antiferromagnetically) below $T_N = 7$ K. Above T_N, the paramagnetic state is thought to be a spin-liquid state described approximately by the Kitaev model[11]. In Fig. 4c, we reproduce data from ref. [23] for κ_{xy}/T versus T in $RuCl_3$. Above 100 K, κ_{xy}/T is vanishingly small. Below 100 K, κ_{xy}/T grows gradually with decreasing T down to 20 K or so (and then drops rapidly as T_N is approached). In the regime between 20 K and 100 K, κ_{xy}/T is well described by calculations for the Kitaev model[11], implying that the κ_{xy} signal in $RuCl_3$ comes from itinerant Majorana fermions—exotic neutral excitations of topological character. This interpretation is supported by the observation[24] of a predicted[11] quantization of the thermal Hall conductivity (at low T when antiferromagnatic order is suppressed by applying a field in the 2D planes). Other spin-liquid candidates, such as volborthite[5] and Ca kapellasite[6], exhibit qualitatively similar $\kappa_{xy}(T)$ (Fig. 4d), suggesting that the gradual growth below about 100 K is a general behaviour.

In Fig. 4c, d, we compare our data on LSCO $p = 0.06$ to the data on $RuCl_3$ and Ca kapellasite, respectively. There is a tantalizing similarity in the gradual growth of $|\kappa_{xy}/T|$ below 100 K or so, but there are some differences. First, whereas κ_{xy} is positive in these two spin-liquid candidates, it is negative in cuprates. (This may reflect the particular topological character of the different states.) Second, the signal in LSCO is approximately 10 to 25 times larger (Fig. 4). Last, in LSCO, κ_{xy}/T continues to grow down to the lowest measured temperature (but it may well drop below about 5–10 K).

In summary, the thermal Hall effect in cuprates reveals a hitherto unknown facet of both the enigmatic pseudogap phase and the Mott insulator, reminiscent of a spin liquid. It points to a state with chirality[7]. It will be interesting to see whether models of topological order[10], spin–charge separation[25] or current loops[26], for example, may be consistent with the giant κ_{xy} signal that appears below p^*. A recent calculation shows that neutral spinons in certain states with topological order on a square lattice can produce a substantially enhanced thermal Hall conductivity[27].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1375-0.

Received: 18 December 2018; Accepted: 26 April 2019;
Published online 17 July 2019.

1. Keimer, B. et al. From quantum matter to high-temperature superconductivity in copper oxides. *Nature* **518**, 179–186 (2015).
2. Proust, C. & Taillefer, L. The remarkable underlying ground states of cuprate superconductors. *Annu. Rev. Condens. Matter Phys.* **10**, 409–429 (2019).
3. Ideue, T. et al. Giant thermal Hall effect in multiferroics. *Nat. Mater.* **16**, 797–802 (2017).
4. Kasahara, Y. et al. Unusual thermal Hall effect in a Kitaev spin liquid candidate α-$RuCl_3$. *Phys. Rev. Lett.* **120**, 217205 (2018).
5. Watanabe, D. et al. Emergence of nontrivial magnetic excitations in a spin-liquid state of kagomé volborthite. *Proc. Natl Acad. Sci. USA* **113**, 8653–8657 (2016).
6. Doki, H. et al. Spin thermal Hall conductivity of a Kagome antiferromagnet. *Phys. Rev. Lett.* **121**, 097203 (2018).
7. Lee, H., Han, J. H. & Lee, P. A. Thermal Hall effect of spins in a paramagnet. *Phys. Rev. B* **91**, 125413 (2015).
8. Monthoux, P., Pines, D. & Lonzarich, G. G. Superconductivity without phonons. *Nature* **450**, 1177–1183 (2007).
9. Kyung, B. et al. Pseudogap induced by short-range spin correlations in a doped Mott insulator. *Phys. Rev. B* **73**, 165114 (2006).
10. Scheurer, M. S. et al. Topological order in the pseudogap metal. *Proc. Natl Acad. Sci. USA* **115**, E3665–E3672 (2018).
11. Nasu, J., Yoshitake, J. & Motome, Y. Thermal transport in the Kitaev model. *Phys. Rev. Lett.* **119**, 127204 (2017).
12. Hirschberger, M. et al. Large thermal Hall conductivity of neutral spin excitations in a frustrated quantum magnet. *Science* **348**, 106–109 (2015).
13. Zhang, Y. et al. Giant enhancement of the thermal Hall conductivity κ_{xy} in the superconductor $YBa_2Cu_3O_7$. *Phys. Rev. Lett.* **86**, 890–893 (2001).
14. Durst, A. C., Vishwanath, A. & Lee, P. A. Weak-field thermal Hall conductivity in the mixed state of d-wave superconductors. *Phys. Rev. Lett.* **90**, 187002 (2003).
15. Cvetkovic, V. & Vafek, O. Berry phases and the intrinsic thermal Hall effect in high-temperature cuprate superconductors. *Nat. Commun.* **6**, 6518 (2015).
16. Grissonnanche, G. et al. Wiedemann-Franz law in the underdoped cuprate superconductor $YBa_2Cu_3O_y$. *Phys. Rev. B* **93**, 064513 (2016).

17. Collignon, C. et al. Fermi-surface transformation across the pseudogap critical point of the cuprate superconductor $La_{1.6-x}Nd_{0.4}Sr_xCuO_4$. *Phys. Rev. B* **95**, 224517 (2017).
18. Kawasaki, S. et al. Carrier-concentration dependence of the pseudogap ground state of superconducting $Bi_2Sr_{2-x}La_xCuO_6$ revealed by ^{63,65}Cu-nuclear magnetic resonance in very high magnetic fields. *Phys. Rev. Lett.* **105**, 137002 (2010).
19. Strohm, C., Rikken, G. L. J. A. & Wyder, P. Phenomenological evidence for the phonon Hall effect. *Phys. Rev. Lett.* **95**, 155901 (2005).
20. Sugii, K. et al. Thermal Hall effect in a spin-phonon glass $Ba_3CuSb_2O_9$. *Phys. Rev. Lett.* **118**, 145902 (2017).
21. Nachumi, B. et al. Muon spin relaxation study of the stripe phase order in $La_{1.6-x}Nd_{0.4}Sr_xCuO_4$ and related 214 cuprates. *Phys. Rev. B* **58**, 8760–8772 (1998).
22. Katsura, H., Nagaosa, N. & Lee, P. A. Theory of the thermal Hall effect in quantum magnets. *Phys. Rev. Lett.* **104**, 066403 (2010).
23. Hentrich, R. et al. Large thermal Hall effect in α-$RuCl_3$: evidence for heat transport by Kitaev-Heisenberg paramagnons. *Phys. Rev. B* **99**, 085136 (2019).
24. Kasahara, Y. et al. Majorana quantization and half-integer thermal quantum Hall effect in a Kitaev spin liquid. *Nature* **559**, 227–231 (2018).
25. Kivelson, S. A., Rokhsar, D. S. & Sethna, J. P. Topology of the resonating valence-bond state: solitons and high-T_c superconductivity. *Phys. Rev. B* **35**, 8865 (1987).
26. Varma, C. M. Theory of the pseudogap state of the cuprates. *Phys. Rev. B* **73**, 155113 (2006).
27. Samajdar, R., Chatterjee, S., Sachdev, S. & Scheurer, M. Thermal Hall effect in square-lattice spin liquids: a Schwinger boson mean-field study. *Phys. Rev. B* **99**, 165126 (2019).
28. Onose, Y. et al. Observation of the magnon Hall effect. *Science* **329**, 297–299 (2010).
29. Cyr-Choinière, O. et al. Pseudogap temperature T^* of cuprate superconductors from the Nernst effect. *Phys. Rev. B* **97**, 064502 (2018).
30. Michon, B. et al. Thermodynamic signatures of quantum criticality in cuprate superconductors. *Nature* **567**, 218–222 (2019).
31. Klauss, H.-H. et al. From antiferromagnetic order to static magnetic stripes: the phase diagram of $(La, Eu)_{2-x}Sr_xCuO_4$. *Phys. Rev. Lett.* **85**, 4590–4593 (2000).
32. Hücker, M. et al. Coupling of stripes to lattice distortions in cuprates and nickelates. *Physica C* **460–462**, 170–173 (2007).

Acknowledgements We thank L. Balents, K. Behnia, S. Chatterjee, B. D. Gaulin, H. J. Han, S. M. Hayden, C. Hess, S. A. Kivelson, H. Y. Kee, P. A. Lee, Y. S. Lee, A. Rosch, S. Sachdev, M. Scheurer, T. Senthil, A.-M. S. Tremblay, C. M. Varma and S. Verret for helpful and stimulating discussions. L.T. acknowledges support from the Canadian Institute for Advanced Research (CIFAR) as a CIFAR Fellow and funding from the Natural Sciences and Engineering Research Council of Canada (NSERC), the Fonds de recherche du Québec–Nature et Technologies (FRQNT), the Canada Foundation for Innovation (CFI), and a Canada Research Chair. This research was undertaken thanks in part to funding from the Canada First Research Excellence Fund. Part of this work was funded by the Gordon and Betty Moore Foundation's EPiQS Initiative (grant GBMF5306 to L.T.). J.-S.Z was supported by NSF MRSEC DMR-1720595 in the US.

Reviewer information *Nature* thanks Kwang-yong Choi, Patrick Lee and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions G.G., A.L., S.B., E.L., V.Z., M.L., F.L., A.G. and N.D.-L. performed the thermal Hall conductivity measurements. G.G., A.L., S.B., E.L., V.Z., M.L. and N.D.-L. performed the electrical Hall conductivity measurements. J.-S.Z. grew the Nd-LSCO single crystals. S.P., T.T. and H.T. grew the Eu-LSCO and LSCO single crystals. S.O. grew the Bi2201 single crystal. G.G., N.D.-L. and L.T. wrote the manuscript, in consultation with all authors. L.T. supervised the project.

Competing interests The authors declare no competing interests.

Additional information
Extended data figures and tables is available for this paper at https://doi.org/10.1038/s41586-019-1375-0.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to G.G. or L.T.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Samples. *Nd-LSCO.* Single crystals of $La_{2-y-x}Nd_ySr_xCuO_4$ (Nd-LSCO) were grown at the University of Texas at Austin using a travelling-float-zone technique, with a Nd content $y = 0.4$ and nominal Sr concentrations $x = 0.20$, 0.21, 0.22, 0.23 and 0.25. The hole concentration p is given by $p = x$, with an error bar ± 0.003, except for the $x = 0.25$ sample, for which the doping is $p = 0.24 \pm 0.005$ (for more details, see ref. [17]). The value of T_c, defined as the point of zero resistance, is $T_c = 15.5$, 15, 14.5, 12 and 11 K for samples with $x = 0.20$, 0.21, 0.22, 0.23 and 0.24, respectively. The pseudogap critical point in Nd-LSCO is at $p^* = 0.23$ (ref. [17]).

Eu-LSCO. Single crystals of $La_{2-y-x}Eu_ySr_xCuO_4$ (Eu-LSCO) were grown at the University of Tokyo using a travelling-float-zone technique, with a Eu content $y = 0.2$ and nominal Sr concentrations $x = 0.08$, 0.21 and 0.24. The hole concentration p is given by $p = x$, with an error bar of ± 0.005. The value of T_c, defined as the point of zero resistance, is $T_c = 3$, 14 and 9 K for samples with $x = 0.08$, 0.21 and 0.24, respectively. The pseudogap critical point in Eu-LSCO is at $p^* = 0.23$ (ref. [30]).

LSCO. Single crystals of $La_{2-x}Sr_xCuO_4$ (LSCO) were grown at the University of Tokyo using a travelling-float-zone technique, with nominal Sr concentrations $x = 0.0$ (that is, La_2CuO_4) and 0.06. The hole concentration p is $p \approx 0$ and $p = 0.06 \pm 0.005$, respectively. The value of T_c, defined as the point of zero resistance, is $T_c = 0$ and 5 K for samples with $x = 0.0$ and 0.06, respectively. The pseudogap critical point in LSCO is at $p^* \approx 0.18$ (ref. [29]).

Bi2201. Our single crystal of $Bi_2Sr_{2-x}La_xCuO_{6+\delta}$ (Bi2201) was grown at CRIEPI in Kanagawa using a travelling-float-zone technique[33], with La content $x = 0.2$. The value of T_c, defined as the onset of the drop in magnetization, is $T_c = 18$ K. Given its x and T_c values, the doping of this overdoped sample is such that $p < p^*$ (ref. [18]).

Transport measurements. Our comparative study of heat and charge transport was performed by measuring the thermal Hall conductivity κ_{xy} and the electrical Hall conductivity σ_{xy} on the same sample, using the same contacts made of silver epoxy H20E annealed at high temperature in oxygen.

Thermal measurements. A constant heat current $\boldsymbol{Q}$ was sent in the basal plane of the single crystal (along x), generating a longitudinal temperature difference $\Delta T_x = T^+ - T^-$ (Fig. 2c). The thermal conductivity along the x axis is given by $\kappa_{xx} = (Q/\Delta T_x)(L/wt)$, where L is the separation (along x) between the two points at which T^+ and T^- are measured, w is the width of the sample (along y) and t its thickness (along z). By applying a magnetic field $\boldsymbol{H}$ along the c axis of the crystal (along z), normal to the CuO_2 planes, one generates a transverse gradient ΔT_y (Fig. 2c). The thermal Hall conductivity is defined as $\kappa_{xy} = -\kappa_{yy}(\Delta T_y/\Delta T_x)(L/w)$, where κ_{yy} is the longitudinal thermal conductivity along the y axis. In this study, we take $\kappa_{yy} = \kappa_{xx}$. The thermal Hall conductivity κ_{xy} of our samples was measured in magnetic fields up to $H = 18$ T. The measurement procedure is described in detail elsewhere[16].

Electrical measurements. The longitudinal resistivity ρ_{xx} and Hall resistivity ρ_{xy} were measured in magnetic fields up to 16 T in a Quantum Design PPMS in Sherbrooke. (For Nd-LSCO $p = 0.20$, σ_{xy} was measured at $H = 33$ T (ref. [17]).) The measurements were performed using a conventional six-point configuration with a current excitation of 2 mA, using the same contacts as for the thermal measurements (Fig. 2c). The electrical Hall conductivity σ_{xy} is given by $\sigma_{xy} = \rho_{xy}/(\rho_{xx}^2 + \rho_{xy}^2)$.

Field dependence of the thermal Hall conductivity. All of the data reported here were taken in a magnetic field (normal to the CuO_2 planes) large enough to fully suppress superconductivity, and thereby access the normal state of Nd-LSCO, Eu-LSCO, LSCO and Bi2201. Indeed, a field of 15 T is sufficient to do this in all samples presented here, down to at least 5 K. In the normal state, κ_{xy} has an intrinsic field dependence. In Extended Data Fig. 4, we show how κ_{xy} in LSCO $p = 0.06$, where $T_c = 5$ K, depends on magnetic field for $T > T_c$: the linear H dependence of κ_{xy} at high T becomes sublinear at low T.

It may be worth pointing out that the sudden appearance of a new negative κ_{xy} signal below p^* is not correlated with any change in the superconducting properties of the sample. The easiest way to see this is to compare Nd-LSCO or Eu-LSCO at $p = 0.24$ and $p = 0.21$. While the superconducting properties at $p = 0.24$ and $p = 0.21$ are very similar—that is, $T_c \approx 10$ K versus 15 K and $H_{c2} \approx 10$ T versus 15 T (ref. [30])—the κ_{xy} response is totally different (at low T): positive at $p = 0.24$, negative at $p = 0.21$ (Fig. 3).

Thermal Hall conductivity in YBCO. In YBCO at $p = 0.11$, there is huge negative κ_{xy} signal in the field-induced normal state[2]. In this excellent metal, whose Fermi surface is reconstructed by charge-density-wave order into a small electron pocket of high mobility[2], the electrical Hall conductivity σ_{xy} is equally huge. In fact, the Wiedemann–Franz law was found to hold, namely $\kappa_{xy}/T = L_0\sigma_{xy}$ as $T \to 0$, within error bars of $\pm 15\%$ (ref. [16]). In other words, the negative κ_{xy} signal in this case is due to the charge carriers (that is, to electrons). However, because the $\pm 15\%$ uncertainty corresponds to ± 12 mW K^{-2} m^{-1} (in 27 T), it is impossible to know whether the κ_{xy} signal in YBCO might also contain a contribution of -2 to -6 mW K^{-2} m^{-1} from neutral excitations (that is, -1 to -3 mW K^{-2} m^{-1} in 15 T; Fig. 1b).

Mott insulator. We can estimate the doping of our LCO sample (La_2CuO_4) by comparing its resistivity with published data. In Extended Data Fig. 6, we compare the resistivity of our LCO sample to published data by Uchida and co-workers[34] on the most stoichiometric sample of La_2CuO_4 they were able to produce, with the highest resistivity. We see that our LCO sample has a similar resistivity, even slightly higher at low temperature. We conclude that p is very close to zero in our sample. In Extended Data Fig. 6, we also compare with data from Komiya and co-workers[35] on a LSCO sample with Sr content $x = 0.01$. We see that our LCO sample's resistivity is larger by several orders of magnitude. We conclude that $p < 0.01$ in our LCO sample.

In Extended Data Fig. 6, we compare the resistivity of our sample of LCO and our sample of LSCO with $p = 0.06$. We see that their resistivities at low T differ by 7–8 orders of magnitude. This shows that although the two samples have very similar κ_{xy} curves (Fig. 1b), they are electrically very different.

Thermal Hall signal from magnons. In undoped La_2CuO_4, magnons have been well characterized by inelastic neutron scattering measurements[36]. There are two magnon branches, each with its own spin gap, of magnitude 26 K and 58 K, respectively. The thermal conductivity of magnons, κ_{mag}, is therefore thermally activated at $T < 26$ K, so that κ_{mag} decreases exponentially at low T. Hess and co-workers have estimated κ_{mag} in La_2CuO_4 by taking the difference between in-plane and out-of-plane conductivities[37]. In Extended Data Fig. 5, we see that κ_{mag}/T decreases monotonically as $T \to 0$ below 150 K.

By contrast, κ_{xy}/T in La_2CuO_4 increases monotonically with decreasing T, all the way down to $T \approx 5$ K (Extended Data Fig. 5), a temperature 5 times smaller than the smallest gap, where there are no thermally excited magnons. Moreover, when we move up in doping to $p = 0.06$, where antiferromagnetic order is gone and LSCO is in a very different magnetic state (Fig. 1a), without well-defined magnons or a spin gap, $\kappa_{xy}(T)$ is essentially identical to that in La_2CuO_4 (Fig. 1b). We conclude that magnons are not responsible for the large negative κ_{xy} in cuprates.

Note, moreover, that a collinear antiferromagnetic order on a square lattice (such as that found in La_2CuO_4) is expected[22] to yield $\kappa_{xy} = 0$. A non-zero κ_{xy} signal could come from the canting of spins out of the CuO_2 planes, but one would expect it to be very small[27]—and it would still vanish at low T because of the gap in the magnon spectrum. Note also that there could be some low-energy spin excitations in La_2CuO_4 besides the well-known magnons. Magnetic susceptibility measurements in La_2CuO_4 and lightly doped LSCO have revealed some unusual features, not consistent with a simple Néel state[38].

Thermal Hall signal from phonons. Phonons can produce a non-zero κ_{xy} signal if they undergo scattering by spins[3,19]. Spin scattering of phonons can be detected through its impact on κ_{xx}. First, it reduces the magnitude of κ_{xx} relative to its value without spin scattering. A good example of this is provided by the insulators $Y_2Ti_2O_7$ and $Tb_2Ti_2O_7$. In non-magnetic $Y_2Ti_2O_7$, $\kappa_{xx}(T)$ is large and typical of phonons in non-magnetic insulators (Extended Data Fig. 2a). In isostructural $Tb_2Ti_2O_7$, which has a large moment on the Tb ion, $\kappa_{xx}(T)$ is massively reduced (Extended Data Fig. 2a), as phonons undergo strong spin scattering. At $T = 15$ K, κ_{xx} is 15 times smaller in $Tb_2Ti_2O_7$.

A second signature of the spin scattering of phonons is a field dependence of κ_{xx}. In $Tb_2Ti_2O_7$, a field of 8 T causes a 30% reduction in κ_{xx} at $T = 15$ K (ref. [12]; Fig. 4a, Extended Data Fig. 2b, Table 1). In the multiferroic material $(Fe,Zn)_2Mo_3O_8$, where the spin–phonon coupling is known to be very strong, a field of 9 T causes a 30% reduction in κ_{xx} at $T = 30$ K (ref. [3]; Fig. 4a, Table 1).

Let us now look for those two signatures in cuprates. First, in Nd-LSCO, where the negative κ_{xy} signal is absent at $p = 0.24$ and present at $p = 0.21$, with a magnitude about 10 times larger than in $Tb_2Ti_2O_7$. If this very large κ_{xy} signal is due to phonons, then there must be some very strong spin scattering of phonons that appears below $p = 0.24$, which would show up as a massive decrease in κ_{xx}. In Extended Data Fig. 3, we see that there is no decrease of κ_{xx} in going from $p = 0.24$ to $p = 0.21$—on the contrary, κ_{xx} increases.

Second, we look at the field dependence of κ_{xx} in LSCO $p = 0.06$, where the negative κ_{xy} signal is about 20 times larger than in $Tb_2Ti_2O_7$, at $T = 15$ K and $H = 8$ T (ref. [12]; Extended Data Fig. 2b, d, Table 1). In LSCO, the change in κ_{xx} induced by a field of 8 T at $T = 14$ K is no more than 1% (Extended Data Figs. 1e and 2d), so about 20 times smaller than in $Tb_2Ti_2O_7$. (This could in part be due to a larger relevant field scale in cuprates.) In addition to being negligible in size, the H dependence of κ_{xx} in LSCO has the wrong T dependence: $[\kappa_{xx}(15\text{ T}) - \kappa_{xx}(1\text{ T})]/T$ drops below 30 K, whereas κ_{xy}/T keeps growing monotonically as $T \to 0$ (Extended Data Fig. 1f).

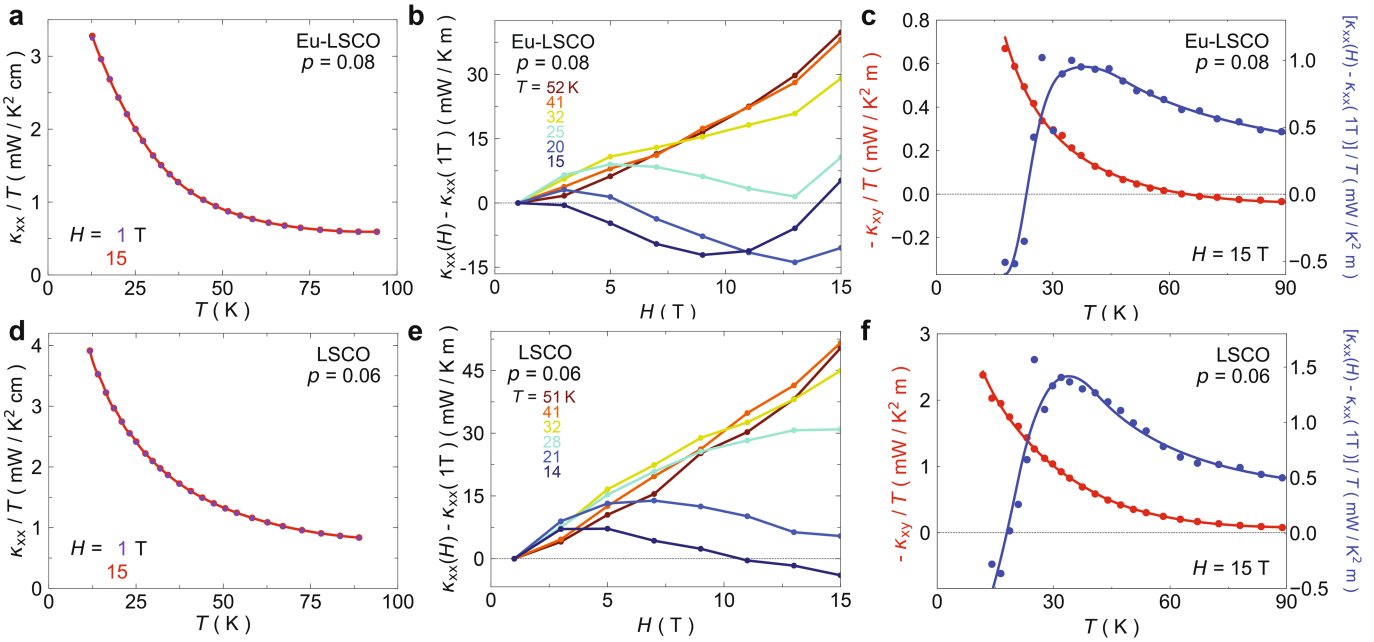
So we find that neither of the two standard signatures of strong phonon–spin scattering is clearly present in cuprates. Moreover, there is no evidence that a new spin state appears below p^* in Nd-LSCO, which would introduce a new mechanism for scattering phonons. On the contrary, static moments present at $p = 0.12$ cease to be detected (by μSR) at $p = 0.20$ (ref. [21]), so that $p = 0.20$ and $p = 0.24$ are equally non-magnetic from the μSR point of view.

We conclude that there is no clear evidence to suggest that phonons are responsible for the large negative κ_{xy} in cuprates that appears below p^*.

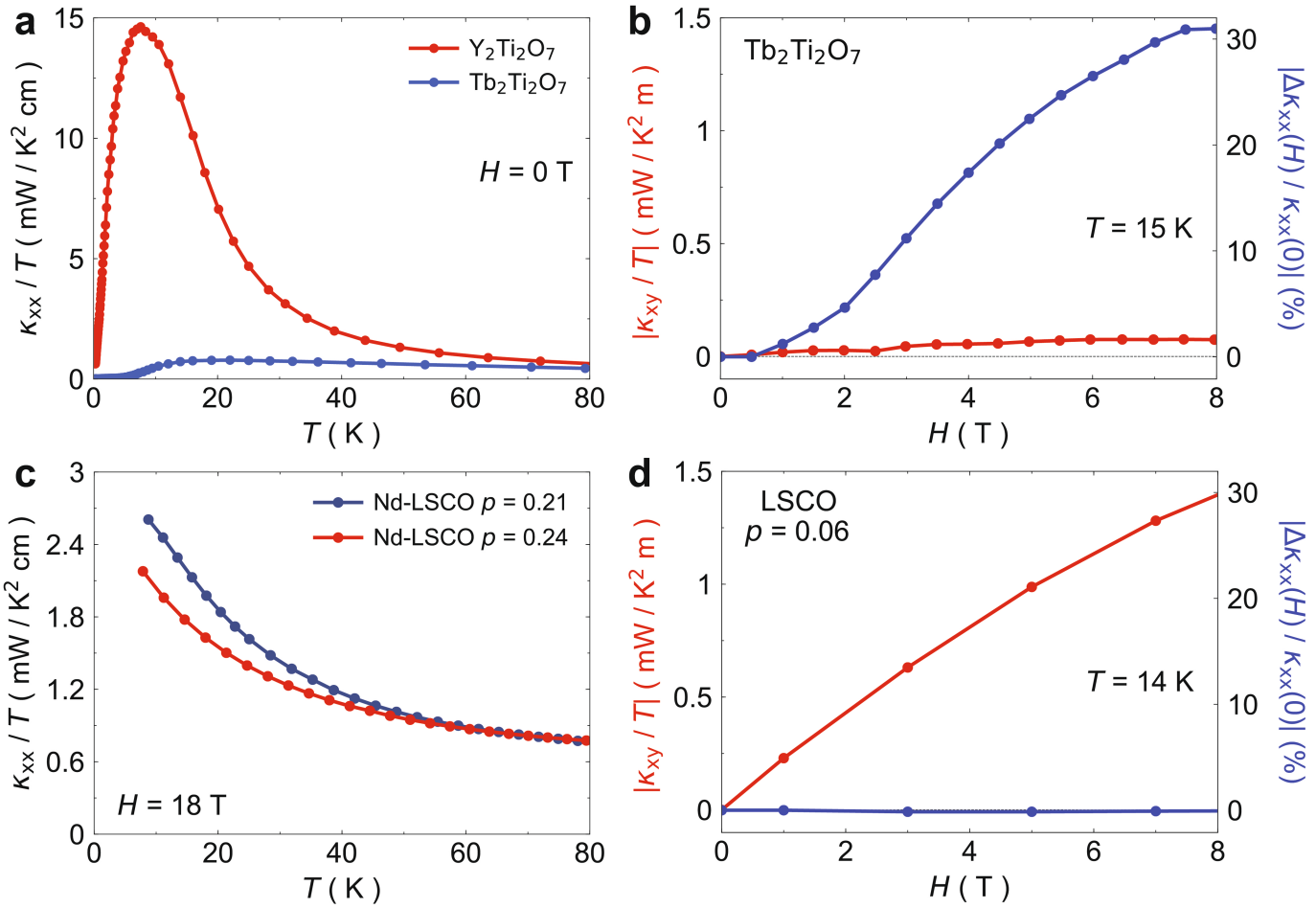
Data availability

The data that support the plots within this paper and other findings of this study are available from the corresponding authors upon reasonable request.

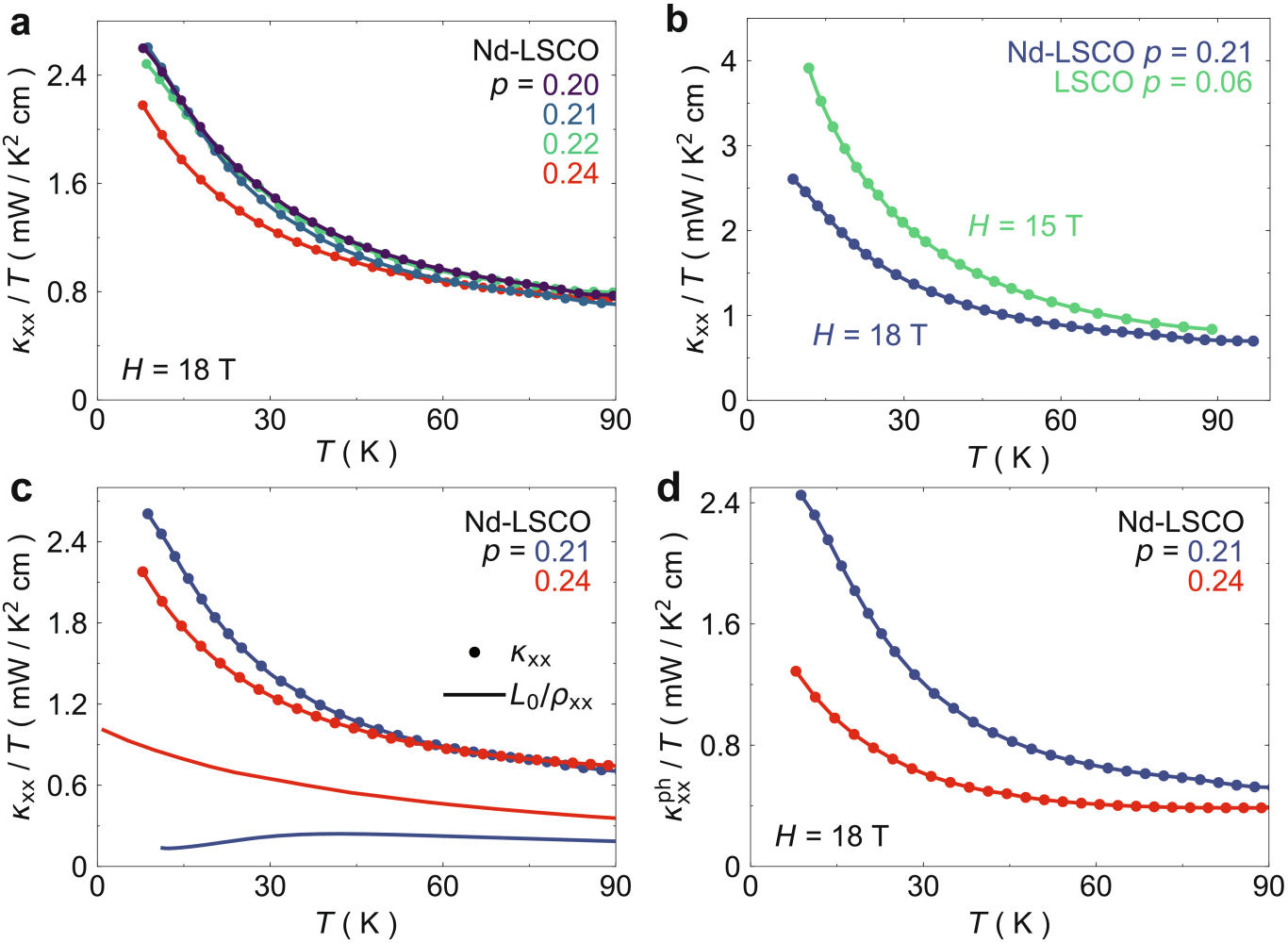
33. Ono, S. & Ando, Y. Evolution of the resistivity anisotropy in $Bi_2Sr_{2-x}La_xCuO_{6+\delta}$ single crystals for a wide range of hole doping. *Phys. Rev. B* **67**, 104512 (2003).
34. Uchida, S. et al. Electric and magnetic properties of La_2CuO_4. *Jpn. J. Appl. Phys.* **26**, L445 (1987).
35. Komiya, S., Ando, Y., Sun, X. F. & Lavrov, A. N. c-axis transport and resistivity anisotropy of lightly to moderately doped $La_{2-x}Sr_xCuO_4$ single crystals: implications on the charge transport mechanism. *Phys. Rev. B* **65**, 214535 (2002).
36. Keimer, B. et al. Soft phonon behavior and magnetism at the low temperature structural phase transition of $La_{1.65}Nd_{0.35}CuO_4$. *Z. Phys. B* **91**, 373–382 (1993).
37. Hess, C. et al. Magnon heat transport in doped La_2CuO_4. *Phys. Rev. Lett.* **90**, 197002 (2003).
38. Lavrov, A. N., Ando, Y., Komiya, S. & Tsukada, I. Unusual magnetic susceptibility anisotropy in untwinned $La_{2-x}Sr_xCuO_4$ single crystals in the lightly doped region. *Phys. Rev. Lett.* **87**, 017007 (2001).
39. Li, Q. J. et al. Phonon-glass-like behavior of magnetic origin in single crystal $Tb_2Ti_2O_7$. *Phys. Rev. B* **87**, 214408 (2013).
40. Michon, B. et al. Wiedemann-Franz law and abrupt change in conductivity across the pseudogap critical point of a cuprate superconductor. *Phys. Rev. X* **8**, 041010 (2018).



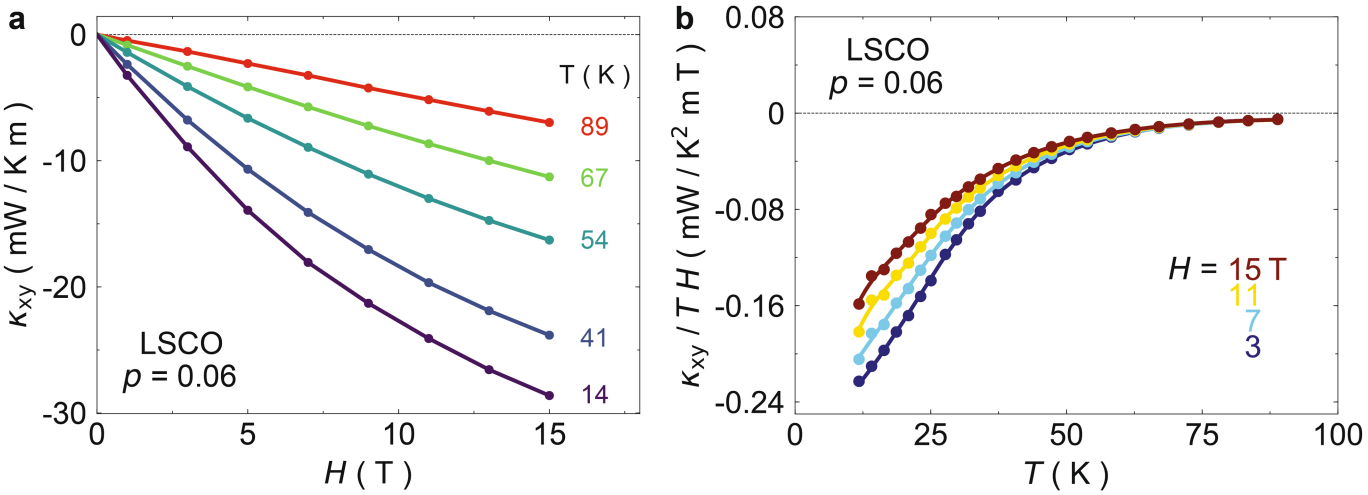
Extended Data Fig. 1 | Magnetic field dependence of κ_{xx}. a–f, Field dependence of κ_{xx} in Eu-LSCO $p = 0.08$ (top panels) and LSCO $p = 0.06$ (bottom panels), displayed in three ways. **a**, **d**, Plot of κ_{xx}/T versus T at $H = 1$ T (blue) and $H = 15$ T (red) (data points). The difference between the two curves is very small, not visible by eye. **b**, **e**, Plot of the change in κ_{xx} with field measured relative to its value at $H = 1$ T, that is, $[\kappa_{xx}(H) - \kappa_{xx}(1\text{ T})]$ versus H, for various temperatures as indicated (data points). **c**, **f**, Change in κ_{xx} between 15 T and 1 T, plotted as $[\kappa_{xx}(H) - \kappa_{xx}(1\text{ T})]/T$ versus T (blue, right axis), compared to $\kappa_{xy}(15\text{ T})/T$ versus T (red, left axis) (data points). Markers represent data and the line is a guide to the eye. Note how at low T the transverse response grows to be as large, if not larger, than the longitudinal response.



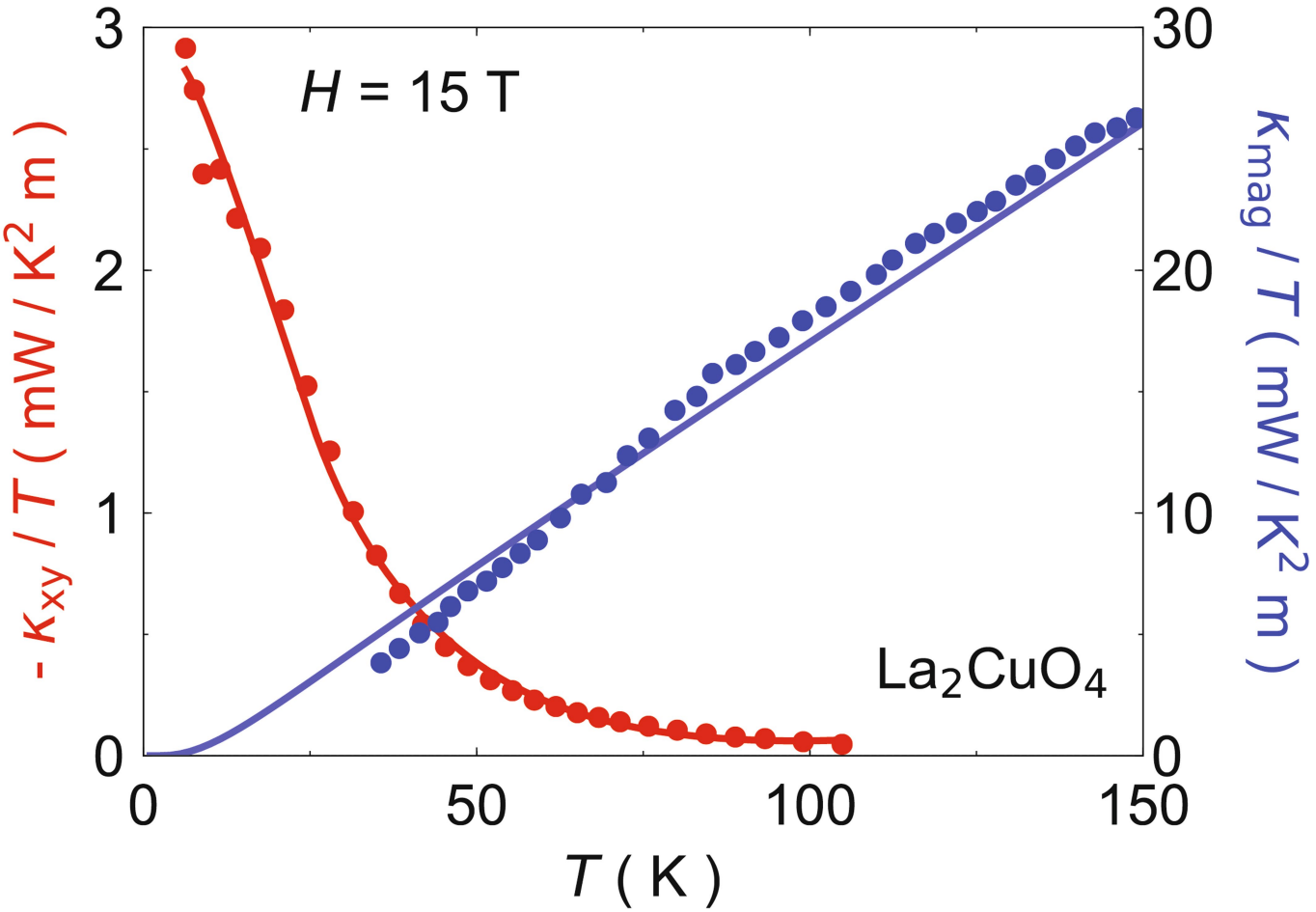
Extended Data Fig. 2 | Comparison of cuprates to other oxides. a, Thermal conductivity of two isostructural oxides, plotted as κ_{xx}/T versus T at $H = 0$, namely $Y_2Ti_2O_7$ (red) and $Tb_2Ti_2O_7$ (blue) (data points[39]). The presence of disordered magnetic moments in $Tb_2Ti_2O_7$ produces a strong scattering of phonons, seen as a massive suppression of κ_{xx} (15-fold at $T = 15$ K). **b**, Field dependence of κ_{xx}, plotted as $\Delta\kappa_{xx}(H)/\kappa_{xx}(0)$ versus H, with $\Delta\kappa_{xx} = \kappa_{xx}(H) - \kappa_{xx}(0)$, at $T = 15$ K (blue data points[12]). The strong effect of field (30% in 8 T) is a direct signature of the strong coupling between phonons and spins in $Tb_2Ti_2O_7$. Also shown is the transverse response in $Tb_2Ti_2O_7$ at $T = 15$ K, plotted as κ_{xy}/T versus H (red data points[12]). Note that in $Y_2Ti_2O_7$, $\kappa_{xy} = 0$ (ref. [12]). **c**, Thermal conductivity of two Nd-LSCO samples, on either side of p^* (red, $p = 0.24$; blue, $p = 0.21$), plotted as κ_{xx}/T versus T at $H = 18$ T (data points). We see that contrary to $Tb_2Ti_2O_7$ (**a**), the appearance of the negative κ_{xy} signal in Nd-LSCO below p^* is not accompanied by a large suppression of κ_{xx} (see Extended Data Fig. 3). **d**, Same as **b** but for LSCO $p = 0.06$, with the same x-axis and y-axis scales and data taken at (nearly) the same temperature (data points). We see that the situation in LSCO is very different to that found in $Tb_2Ti_2O_7$ (**b**): instead of having a small κ_{xy} and a large $\Delta\kappa_{xx}$ (**b**), we now have a large κ_{xy} and a small $\Delta\kappa_{xx}$. Quantitatively, $\kappa_{xy}/\Delta\kappa_{xx} \approx 1$ in LSCO and approximately 0.01 in $Tb_2Ti_2O_7$, at $T = 15$ K and $H = 8$ T (Table 1).



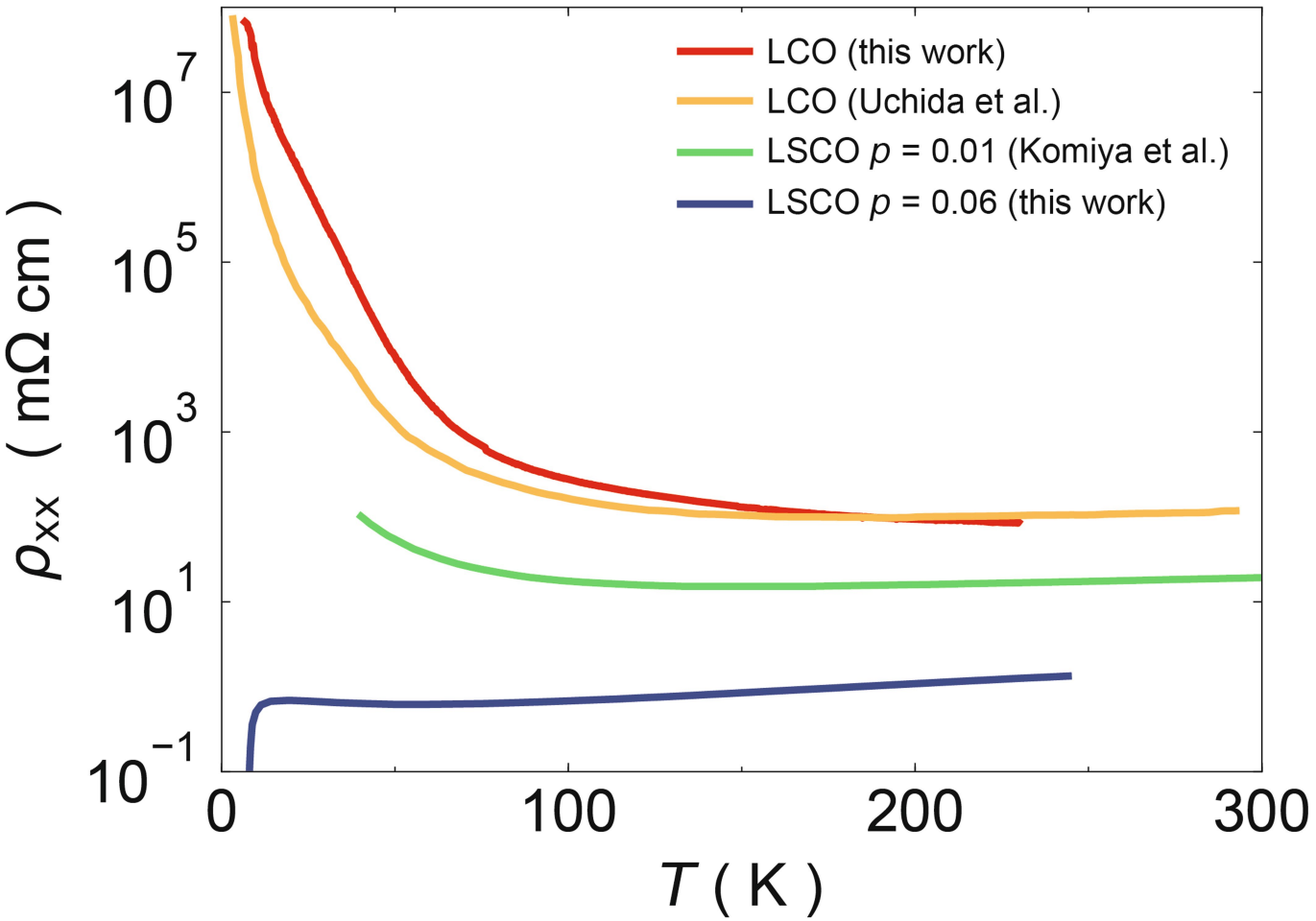
Extended Data Fig. 3 | Change in phonon κ_{xx} across p^* in Nd-LSCO. a, Thermal conductivity of Nd-LSCO at four different dopings, above p^* ($p = 0.24$) and below p^* ($p = 0.20, 0.21, 0.22$), plotted as κ_{xx}/T versus T, at $H = 18$ T (data points). We see that κ_{xx} increases below p^*. **b**, Same as **a** but for Nd-LSCO $p = 0.21$ (blue; $H = 18$ T) and LSCO $p = 0.06$ (green, $H = 16$ T). We see that κ_{xx} continues to increase as we lower p further. This shows that phonons conduct better at lower p. A natural explanation is that they are less scattered by charge carriers as the material becomes less metallic. **c**, Same data as in **a** for Nd-LSCO $p = 0.21$ (blue data points) and $p = 0.24$ (red data points), compared to the electrical conductivity of those same samples, plotted as L_0/ρ versus T (lines; measured at $H = 33$ T (ref. [17])). The latter curves are a reasonable estimate of the electronic thermal conductivity κ_{xx}^{el}, exact at $T \to 0$ (since the Wiedemann–Franz law is satisfied[40]), as seen in Fig. 2a. **d**, Estimate of the phonon conductivity, defined as $\kappa_{xx}^{ph} = \kappa_{xx} - L_0 T/\rho$, plotted as κ_{xx}^{ph}/T versus T (using data from **c**) (data points). We see that $\kappa_{xx}^{ph}(T)$ increases upon crossing below p^*, most probably because electron–phonon scattering is weakened by the loss of carrier density. There is no evidence that the phonons suddenly suffer from the onset of strong spin scattering below p^* (which would cause $\kappa_{xx}^{ph}(T)$ to drop below p^*), such as would be required to explain the appearance of the large negative κ_{xy} signal below p^* (Fig. 3) as being due to phonon transport.



Extended Data Fig. 4 | Magnetic field dependence of κ_{xy} in LSCO. **a**, Field dependence of the thermal Hall conductivity of LSCO at $p = 0.06$, plotted as κ_{xy} versus H at various temperatures, as indicated (data points). The dependence of κ_{xy} on H is linear at high T and it becomes sublinear at lower T. **b**, Deviation from linearity displayed by plotting $\kappa_{xy}/(TH)$ versus T at four different fields H, as indicated (data points).



Extended Data Fig. 5 | Magnon thermal conductivity in La_2CuO_4. Thermal conductivity of magnons in La_2CuO_4, plotted as κ_{mag}/T versus T (blue data points, right axis; ref. [37]). The solid blue line is a fit to the data using the standard calculation for two magnon branches in 2D, with gaps as measured by neutron inelastic scattering[36], namely $\Delta_1 = 26$ K and $\Delta_2 = 58$ K. Below $T \approx 5$ K, thermally excited magnons are exponentially rare and $\kappa_{mag}/T \approx 0$. In sharp contrast, the thermal Hall conductivity of La_2CuO_4, $|\kappa_{xy}/T|$ (red data points, left axis; the red line is a guide to the eye; Fig. 1b), is largest at $T \approx 5$ K. This comparison shows that the κ_{xy} signal in La_2CuO_4 cannot come from magnon transport.



Extended Data Fig. 6 | Electrical resistivity in La_2CuO_4. Electrical resistivity, ρ_{xx}, of two of our samples—La_2CuO_4 (LCO, red) and LSCO at $p = 0.06$ (blue)—compared with published data for La_2CuO_4 (yellow[34]) and LSCO at $p = 0.01$ (green[35]). This shows that our LCO sample is very close to the Mott insulator La_2CuO_4, being more insulating than LSCO with $p = 0.01$ and much more insulating than our LSCO sample with $p = 0.06$.

LETTER

https://doi.org/10.1038/s41586-019-1388-8

Designing minimal and scalable insect-inspired multi-locomotion millirobots

Zhenishbek Zhakypov[1], Kazuaki Mori[2], Koh Hosoda[2] & Jamie Paik[1]*

In ant colonies, collectivity enables division of labour and resources[1–3] with great scalability. Beyond their intricate social behaviours, individuals of the genus *Odontomachus*[4], also known as trap-jaw ants, have developed remarkable multi-locomotion mechanisms to 'escape-jump' upwards when threatened, using the sudden snapping of their mandibles[5], and to negotiate obstacles by leaping forwards using their legs[6]. Emulating such diverse insect biomechanics and studying collective behaviours in a variety of environments may lead to the development of multi-locomotion robotic collectives deployable in situations such as emergency relief, exploration and monitoring[7]; however, reproducing these abilities in small-scale robotic systems with simple design and scalability remains a key challenge. Existing robotic collectives[8–12] are confined to two-dimensional surfaces owing to limited locomotion, and individual multi-locomotion robots[13–17] are difficult to scale up to large groups owing to the increased complexity, size and cost of hardware designs, which hinder mass production. Here we demonstrate an autonomous multi-locomotion insect-scale robot (millirobot) inspired by trap-jaw ants that addresses the design and scalability challenges of small-scale terrestrial robots. The robot's compact locomotion mechanism is constructed with minimal components and assembly steps, has tunable power requirements, and realizes five distinct gaits: vertical jumping for height, horizontal jumping for distance, somersault jumping to clear obstacles, walking on textured terrain and crawling on flat surfaces. The untethered, battery-powered millirobot can selectively switch gaits to traverse diverse terrain types, and groups of millirobots can operate collectively to manipulate objects and overcome obstacles. We constructed the ten-gram palm-sized prototype—the smallest and lightest self-contained multi-locomotion robot reported so far—by folding a quasi-two-dimensional metamaterial[18] sandwich formed of easily integrated mechanical, material and electronic layers, which will enable assembly-free mass-manufacturing of robots with high task efficiency, flexibility and disposability.

The jaw jump and leg jump multi-locomotion mechanisms that have evolved in trap-jaw ants are vital for traversing obstacles that are orders of magnitude larger than their millimetre-sized bodies, avoiding predators and covering large areas in search of food[19] (Extended Data Fig. 1). Engineering the ability to negotiate diverse terrain types at the meso-scale with design scalability remains a major challenge for the hardware design of locomotion mechanisms[20]. Some locomotion strategies, like jumping, necessitate considerable mechanical power[21] to achieve high take-off velocity, whereas walking requires relatively low power, and combining them into a compact robotic body with a minimal but tunable actuation power mechanism suitable for mass manufacturing is difficult. Existing small-scale multi-locomotion robots[13–17] possess individual mechanisms for each locomotion gait, with the associated increase in the number of gear trains, joints and links, which makes manufacturing difficult, and some require external electromagnetic actuation[22,23]. Neither of these approaches offers a compact, scalable and autonomous multi-locomotion robot platform with capabilities similar to those that exist in the natural world.

We report the development of a multi-locomotion origami millirobot called Tribot (Fig. 1a), that addresses the multi-terrain mobility and scalability challenges of small-scale robots using a single, but versatile, locomotion mechanism. Tribot is a three-legged robot with dimensions of 30 mm (width), 58 mm (length) and 58 mm (height) with a Y-shaped flexure hinge (Y-hinge) at the centre, which can open and instantly close its legs by selective activation of three linear spring-type shape-memory alloy (SMA) actuators that function as 'muscles'. Similar to the mandibles of the trap-jaw ant, the Y-hinge forms the basis of the snap-through mechanism that enables Tribot to leap and clear obstacles (Fig. 1b). When the Y-hinge is opened on any of the three sides by a pair of extensor SMA spring actuators to an angle slightly above 180° and then compressed uniaxially by a flexor SMA spring actuator, it experiences instability and 'snaps through' to the side of the applied compressive force with a variable speed proportional to the applied force (Fig. 1c). If the snap occurs at the hinge bottom, Tribot leaps vertically upwards in a height jump, similar to a jaw jump; if the snap is at any of the two hinge sides, the robot leaps horizontally in a distance jump (a leg-jump motion), which is beneficial for striding across gaps. Tribot can also combine both movements in a somersault jump for clearing barrier obstacles (Fig. 1d). In this case, the bottom SMA spring actuator activates shortly before triggering the side spring actuator that snaps the mechanism, so the robot leaps both vertically and horizontally in the air, flipping before landing. To use the same mechanism to enable the robot to walk with periodic short steps over textured terrains, we developed a 'flic-flac' locomotion strategy (also known as a forward-flip or handspring) similar to that used by the Moroccan spider *Cebrennus rechenbergi*[24]. Here, the actuator activation sequence is the same as for the somersault jump, except that the compression of the trigger actuator occurs at a low power so that the robot slightly hops and flips onto the next two legs. This manoeuvre can be produced multiple times, beginning from any of the robot's edges. To achieve transport with fine steps on flat terrain, we incorporated a crawling strategy similar to that used by inchworms. We used the continuous bending ability of the Y-hinge at three sides combined with stick-slip rubber pads attached to the latches on two sides of the robot's legs, that grip and release the ground contact. We embedded two torsional-sheet SMA actuators[25] with micro-heater layers into the latch folds, which change the angle of latches to produce controllable crawling in both the backward and forward directions.

To achieve scalability of the millirobot for collective applications, we combine the automated printed circuit board (PCB) assembly process and the flexibility of smart composite microstructure design[26] to facilitate the integration of the mechanical, material and electronic layers of the robot. This is achieved by processing the layers in two dimensions, laminating them layer by layer and assembling them into three dimensions by folding[27,28] (Fig. 1e). The PCB layer serves as structural backing and for robot autonomy by embedding off-the-shelf electronic components (including a microcontroller, distance and communication sensors and rechargeable batteries), and a Kapton polyimide layer

[1]Reconfigurable Robotics Laboratory, Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland. [2]Adaptive Robotics Laboratory, Osaka University, Osaka, Japan.
*e-mail: jamie.paik@epfl.ch

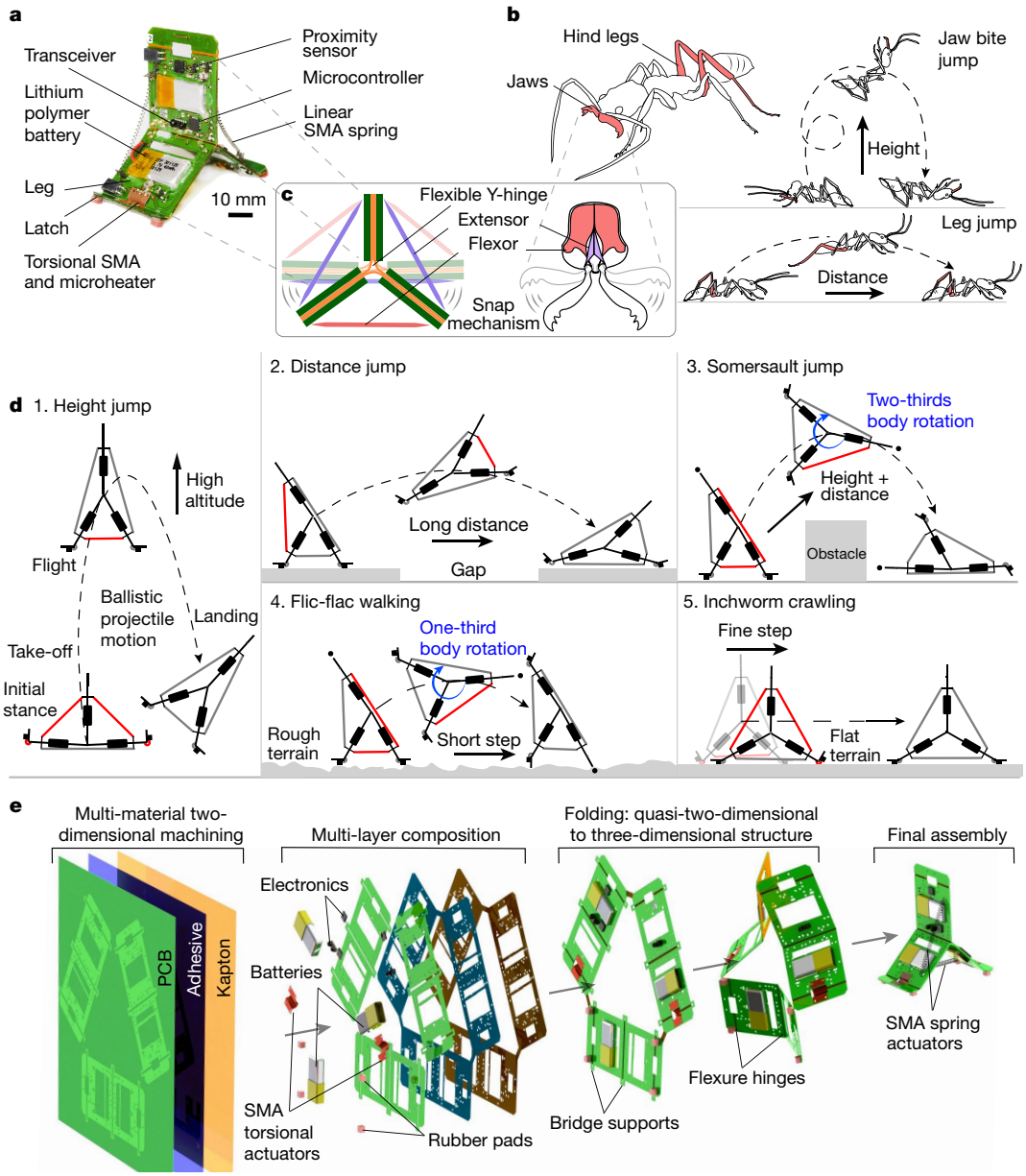


Fig. 1 | Design and fabrication of the trap-jaw-ant-inspired Tribot multi-locomotion millirobot. a, The untethered millirobot Tribot with a Y-hinge controlled by SMA actuators. **b**, The trap-jaw ant uses the snap of its mandible and its hind legs for jumping. **c**, The Y-hinge that connects the three legs 'snaps through' when compressed uniaxially with high force, and bends at low forces and angles less than 180°. **d**, Selective snapping and bending of the Y-hinge generates five locomotion gaits: height (jaw) jumping, distance (leg) jumping, somersault jumping, flic-flac walking and inchworm crawling. The activation pattern is shown by the red-highlighted springs. **e**, The multilayer two-dimesional rapid fabrication and folding assembly process of Tribot.

forms the flexure hinges. The integrated design of the actuators, mechanisms and surface-mounted electronics enables miniaturization of the robot; however, for the current version of Tribot, the major factor determining its size is the capacity of the off-the-shelf battery (3.7 V, 40 mA h), which occupies almost half of the robot's PCB surface area and 40% of its body mass (Extended Data Table 1).

To validate the efficacy and repeatability of Tribot's locomotion gaits, we conducted twelve original locomotion experiments: eleven independent gait tasks across the five gaits: height jumping, distance jumping, somersault jumping, walking and crawling, for various terrain, power and load conditions, each repeated six times, and one continuous 'parkour' (obstacle course) scenario employing multiple gaits, with smooth and rough terrain and an obstacle (Fig. 2, Extended Data Table 2). We studied the robot's motion by recording each experiment on camera at a high frame rate of 250 frames per second (fps) for jumping and in real time (25 fps) for the walking and crawling gaits, and tracking the central Y-hinge using video analysis software. We assessed the robot's vertical leaping capacity by studying height jumps on a flat surface, from its edges with and without the latches with rubber pads (Fig. 2a, Supplementary Video 1). For a trigger Joule heating power of 3.7 W to the flexor SMA spring actuator, Tribot jumps to a height of 140.6 mm on average (almost 2.5 times the robot's height) from the edge without latches, owing to the minimal friction during take-off, and to a height of 72.5 mm from the edge with latches (Fig. 3a). We studied the robot's horizontal distance jump for a trigger power of 3.7 W (Fig. 2b) and 2.7 W, and with an added payload of 5 g (more than 50% the robot's body mass) at a trigger power of 3.7 W; the average jumping distance was 230 mm (almost four times its body length), 140 mm and 110 mm for these tests, respectively. The somersault jump gait was tested with a trigger power of 3.7 W; the average height and distance travelled were 88 mm and 100 mm, respectively, with an average two-thirds body rotation around its central axis in

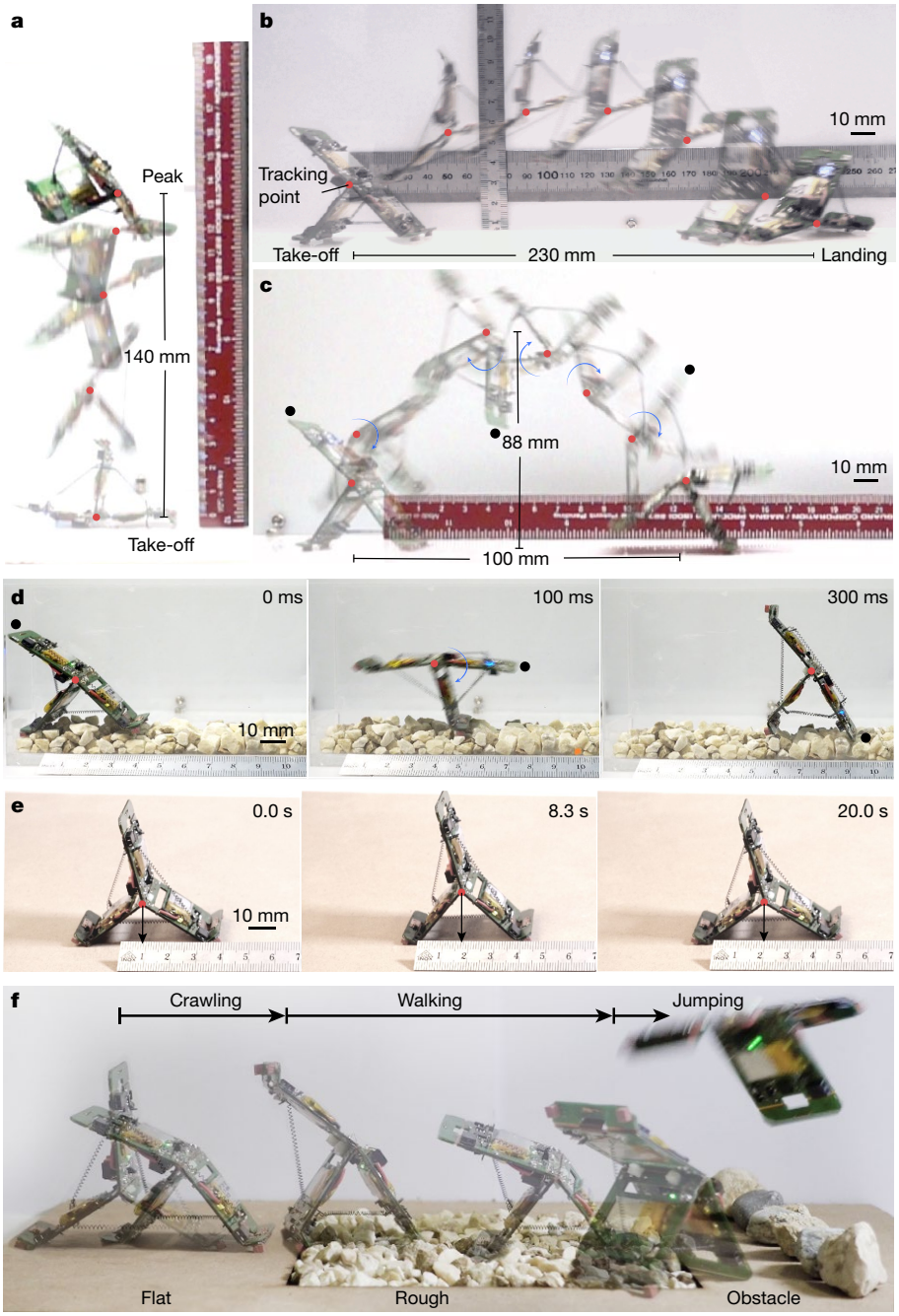


Fig. 2 | Individual multi-locomotion experiments and their combination in the parkour scenario. a, Tribot in height-jump gait on the non-latched side, from take-off to an average peak height of 140 mm. **b**, Tribot in distance-jump gait, from its initial stance to its landing position; 230 mm was the mean jump distance for a power input of 3.7 W. **c**, Tribot performing a somersault jump manoeuvre, during which it rotates in the air before landing; 88 mm was the mean jump height, and 100 mm the mean horizontal jump distance. **d**, Tribot performing a single flic-flac walking manoeuvre on rough terrain with raisin-sized stones. **e**, The robot performing an inchworm crawling step using stick-slip motion on a smooth flat surface. **f**, Demonstration of the adaptability of Tribot's gait in a parkour setup comprising flat terrain, rough terrain and an obstacle. Tribot crawls through the flat section, walks on the rough terrain and somersaults over the obstacle.

the air. To test the effectiveness of the walking gait on flat and rough surfaces and of climbing up a flat slope of 10°, we instructed Tribot via remote control to perform three flic-flac manoeuvres on each terrain, such that it arrived back on its initial edge. To prevent the robot from deviating sideways, we placed it inside a transparent acrylic glass channel slightly wider than its body, with and without raisin-sized stones on the floor (Fig. 2d, Supplementary Video 2). Tribot completes an average repeatable step of 48.8 mm on the flat surface with a small 1.7 W trigger power, but the steps on the textured surface were not repeatable and considerably smaller, at 31.1 mm (Fig. 3b). This is probably due to minimal leg–ground contact with the textured surface, especially on the edges without latches, which causes the legs to slip with each rotation. The average step on the slope was 28.8 mm and required about 30% more power to perform a full flip. We validated the robot's crawling locomotion on a flat surface and on a 10° slope (Fig. 2e, Supplementary Video 3). The robot was programmed to execute six consecutive crawling steps; the mean step size was 4.85 mm on the flat surface and just 2.61 mm on the slope, owing to slippage (Fig. 3c).

We assessed the generated mechanical power and energy cost of transport (COT) for all five gaits and compared Tribot's distance-jumping COT to those of existing small-scale multi-locomotion robots and insects (Fig. 3e, Extended Data Table 3). Tribot is the smallest and

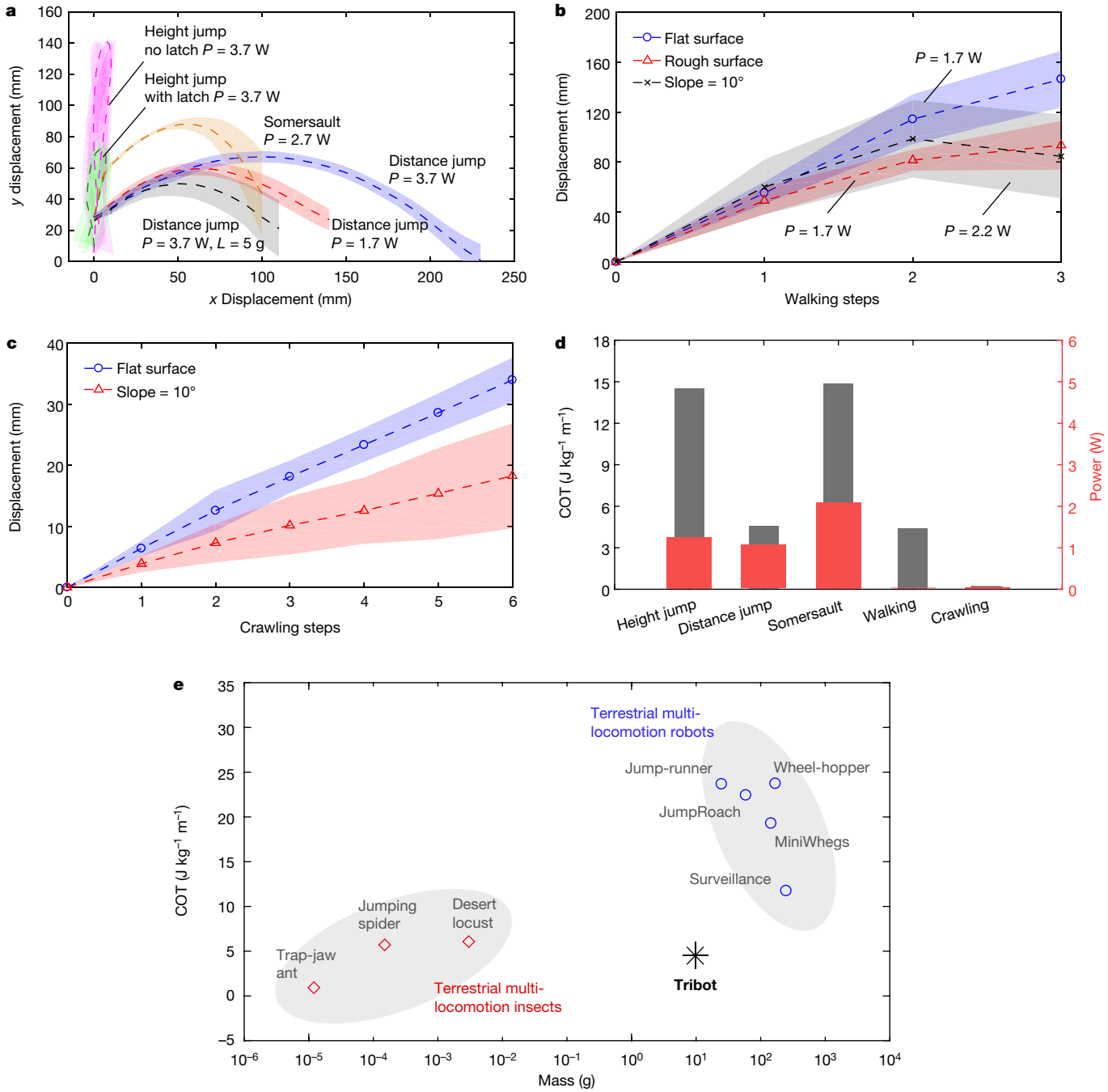


Fig. 3 | Locomotion performance of Tribot in different conditions and its COT compared to robots and insects. a, Tribot's motion projectiles for height, distance and somersault jumping for various SMA flexor spring Joule heating power inputs, *P*, and payloads, *L*, extracted from videos recorded at 250 fps. **b**, Tribot's walking displacement per walking step on flat smooth terrain, flat rough terrain and a 10° smooth slope. The robot flips three times, returning to its initial orientation. **c**, Tribot's inchworm crawling displacement on a flat smooth surface and a smooth 10° slope, measured for six consecutive crawling step cycles. For **a**–**c**, the dashed lines and shaded regions indicate the mean and 1σ, respectively, and each experiment is repeated six times. **d**, The mechanical power output and cost of transport for the five gaits. **e**, Tribot is smaller, lighter and has considerably lower COT for distance jumps than existing small-scale robots[13–17]; its characteristics are comparable to those of insects[5,29,30].

lightest among the engineered systems surveyed and has substantially lower COT for distance jumps, comparable to that of trap-jaw ants[5], the desert locust[29] and jumping spiders[30].

To demonstrate Tribot's multi-locomotion abilities on diverse terrain types and against an obstacle, we set up a parkour experiment (Fig. 2f). Owing to the difficulty in predicting the locomotion manoeuvre outcomes on rough terrain, we controlled the walking and jumping phases remotely, whereas the crawling locomotion was pre-programmed. In this test, Tribot crawled through the smooth section until it reached the rough terrain, where it switched to its walking gait. After four flips, Tribot arrived at the obstacle and jumped over it (Fig. 2f, Supplementary Video 4). We observed discrepancy in the walking phase due to poor leg–ground contact that caused slippage, and the robot once landed on the wrong edge. However, because walking is possible on any edge, the robot was able to continue the manoeuvre, demonstrating its versatility.

To display task efficacy in a scalable collective of multi-locomotion millirobots, we set up two experiments, a division-of-labour scenario and a tandem-running scenario (Fig. 4, Supplementary Video 5), both executed autonomously. The first task was to move to a set position a prismatic object (a rectangular block) that was light enough for two robots to push but too heavy for one to move alone (Fig. 4a). Such a simple task, however, required five millirobots: two workers to push

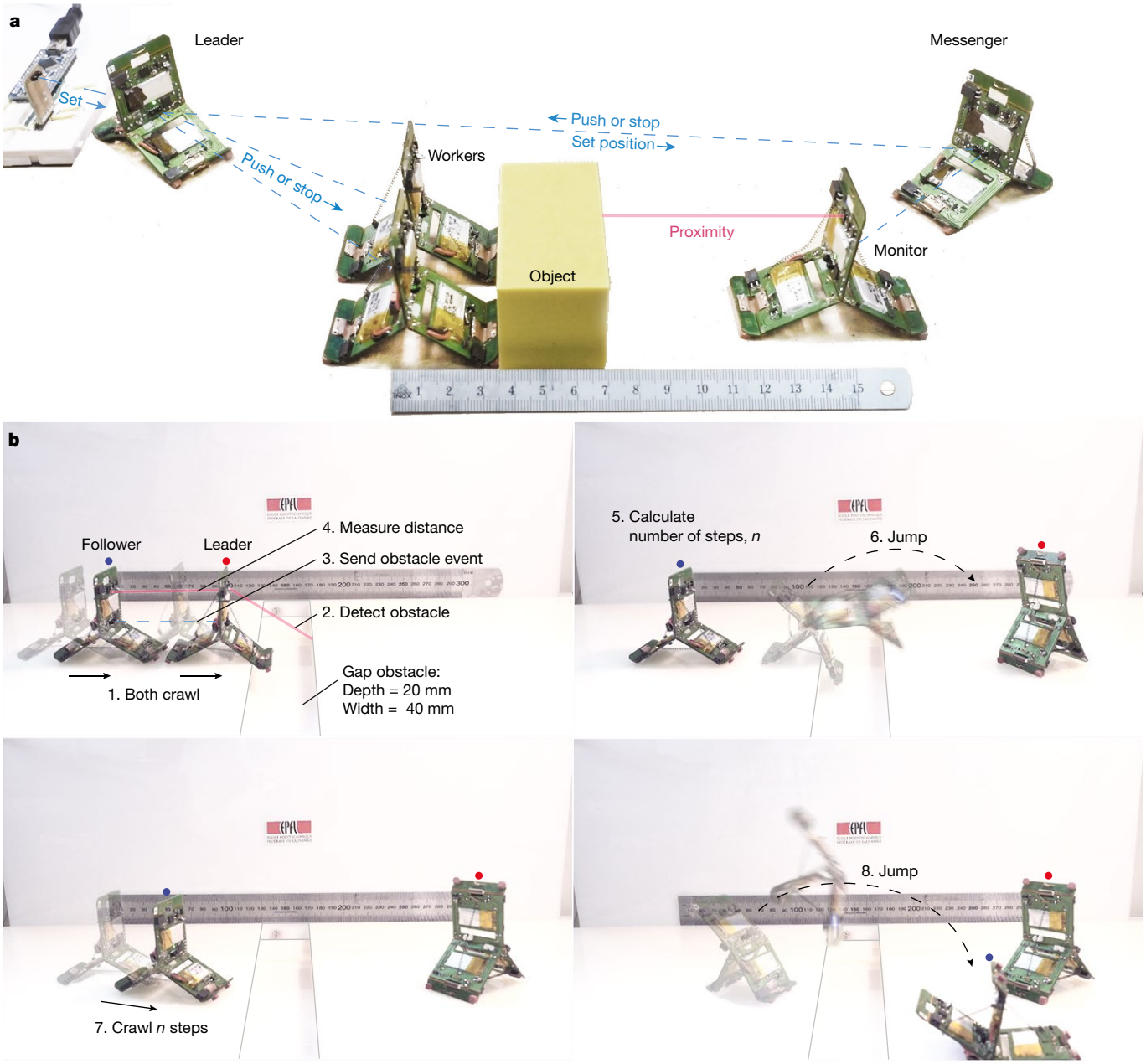


Fig. 4 | Collective labour experiments. a, Division-of-labour experiment. A rectagular prism was pushed to a desired position, using coordination between a leader, two workers, a monitor and a messenger robot. The leader orders two worker robots to push the object while the monitor measures the relative position of the object. As the object interrupts the two-way link between the leader and the monitor, the messenger maintains the communication link. **b**, Tandem-running experiment. Both robots crawl in a line (1) until the leader detects the gap obstacle (2) and transmits the information to the follower (3). Unlike the leader, the follower does not search for obstacles, and therefore saves energy. The follower measures the distance relative to the leader once (4) and calculates the total number of steps, *n*, to crawl to the obstacle (5), while the leader jumps over the obstacle (6). The worker crawls *n* steps (7) before safely jumping across the gap (8) (Supplementary Video 5).

the object towards a monitor, who measured the relative closeness of the object using its proximity sensor and then informed the leader, who in turn coordinated the workers to continue or stop pushing. As the location of the prism interrupted the local communication between the leader and the monitor, the scheme required a messenger, who exchanged information between the leader and monitor via a path that avoided the prism. The data from the proximity sensor of the monitor robot show that the object travelled from a distance of 80 mm from the monitor robot to the programmed distance of 50 mm (Extended Data Fig. 3a). This experiment demonstrates the importance of millirobot scalability for effectively allocating tasks, manipulating objects and resolving communication issues with increased robot population.

To highlight the importance of robot multi-locomotion in collective tasks, we set up a tandem-running experiment with obstacle avoidance that fully demonstrates this functionality in Tribot. In the experiment, two robots, a leader and a follower (Fig. 4b), operate autonomously. The robots crawl in a line, with a fixed step size. The leader, while crawling, continuously scans for obstacles using its proximity sensor whereas the follower only checks messages from the leader, without energy expenditure for scanning. Both robots crawl until the leader detects an obstacle, upon which it conveys this information to the follower (Extended Data Fig. 3b). We chose a wide (40 mm) and deep (20 mm) gap as the obstacle, which may be crossed only by jumping. We demonstrate that the leader, when stopped at the gap edge and tilted forward, notifies the follower of the obstacle. Upon receiving the message, the

follower measures its distance relative to the leader only once and calculates the total number of steps to crawl by dividing that distance by the step size. Once the message is transmitted, the leader jumps over the gap, and the follower subsequently crawls and jumps, following its calculations, without falling in. This demonstration shows not only the expanded capabilities of multi-locomotion millirobot collectives, but also the benefits of collectivity in negotiating obstacles, such as allocating demanding tasks to a leader. Such millirobot tandem-running experiments can be used to further enhance the abilities of collectives of robots to navigate various terrain types and obstacles to effectively plan locomotion through teaching and learning.

We have presented an insect-scale origami robot with a minimal and scalable design that realizes multi-locomotion. The implementations that we present here are expected to facilitate future research into the effect of multi-locomotion ability on the collective behaviours, colony size and task distribution of social insects, that in turn will stimulate the development of algorithms for large-scale collectives of robots with expanded capabilities. The use of a customizable and mass-producible hardware platform forwards investigation into a variety of insect-inspired bio- and neuro-mechanics for millirobots. Although the current version of this hardware has limited manoeuvrability and sensing and computational capacity, it demonstrates applicability to real-world problems, such as emergency mitigation, environmental monitoring and exploration. We now aim to investigate comprehensive design methods[28] and automated fabrication processes to enable on-demand, 'push-button-manufactured' robots and mechanisms accessible to diverse research communities.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1388-8.

Received: 31 August 2018; Accepted: 7 May 2019;
Published online 10 July 2019.

1. Heyman, Y., Shental, N., Brandis, A., Hefetz, A. & Feinerman, O. Ants regulate colony spatial organization using multiple chemical road-signs. *Nat. Commun.* **8**, 15414 (2017).
2. Gordon, D. M. The ecology of collective behavior. *PLoS Biol.* **12**, e1001805 (2014).
3. Franks, N. R. & Richardson, T. Teaching in tandem-running ants. *Nature* **439**, 153 (2006).
4. Sorger, D. M. & Zettel, H. On the ants (Hymenoptera: Formicidae) of the Philippine Islands: V. The genus *Odontomachus Latreille,* 1804. *Myrmecol. News* **14**, 141–163 (2011).
5. Patek, S. N., Baio, J. E., Fisher, B. L. & Suarez, A. V. Multifunctionality and mechanical origins: ballistic jaw propulsion in trap-jaw ants. *Proc. Natl Acad. Sci. USA* **103**, 12787–12792 (2006).
6. Sorger, D. M. Snap! Trap-jaw ants in Borneo also jump using their legs. *Front. Ecol. Environ.* **13**, 574–575 (2015).
7. Fearing, R. S. Challenges for effective millirobots. In *IEEE Int. Sym. Micro-NanoMech. Hum. Sci.* 1–5 (IEEE, 2006).
8. Rubenstein, M., Cornejo, A. & Nagpal, R. Programmable self-assembly in a thousand-robot swarm. *Science* **345**, 795–799 (2014).
9. Kernbach, S. et al. Adaptive collective decision-making in limited robot swarms without communication. *Int. J. Robot. Res.* **32**, 35–55 (2013).
10. Arvin, F., Murray, J., Zhang, C. & Yue, S. Colias: an autonomous micro robot for swarm robotic applications. *Int. J. Adv. Robot. Syst.* **11**, 113 (2014).
11. Pickem, D. et al. The Robotarium: a remotely accessible swarm robotics research testbed. In *2017 IEEE Int. Conf. Rob. Autom.* 1699–1706 (IEEE, 2016).
12. Weston-Dawkes, W. P., Ong, A. C., Majit, M. R. A., Joseph, F. & Tolley, M. T. Towards rapid mechanical customization of cm-scale self-folding agents. In *IEEE Int. Conf. Intel. Rob. Sys.* 4312–4318 (IEEE, 2017).
13. Jung, G.P., Casarez, C. S., Jung, S. P., Fearing, R. S. & Cho, K. J. An integrated jumping–crawling robot using height-adjustable jumping module. In *2016 IEEE Int. Conf. Rob. Autom.* 4680–4685 (IEEE, 2016).
14. Zhao, J., Yan, W., Xi, N., Mutka, M. W. & Xiao, L. A miniature 25 grams running and jumping robot. In *2014 IEEE Int. Conf. Rob. Autom.* 5115–5120 (IEEE, 2014).
15. Morrey, J. M., Lambrecht, B., Horchler, A. D., Ritzmann, R. E. & Quinn, R. D. Highly mobile and robust small quadruped robots. In *2003 IEEE/RSJ Int. Conf. Intel. Rob. Sys.* **1**, 82–87 (IEEE, 2003).
16. Song, G., Yin, K., Zhou, Y. & Cheng, X. A surveillance robot with hopping capabilities for home security. *IEEE Trans. Consum. Electron.* **55**, 2034–2039 (2009).
17. Zhang, Y., Zhang, L., Wang, W., Li, Y. & Zhang, Q. Design and implementation of a two-wheel and hopping robot with a linkage mechanism. *IEEE Access* **6**, 42422–42430 (2018).
18. Bertoldi, K., Vitelli, V., Christensen, J. & van Hecke, M. Flexible mechanical metamaterials. *Nat. Rev. Mater.* **2**, 17066 (2017).
19. Larabee, F. J. & Suarez, A. V. Mandible-powered escape jumps in trap-jaw ants increase survival rates during predator–prey encounters. *PLoS One* **10**, e0124871 (2015).
20. Bergbreiter, S. Effective and efficient locomotion for millimeter-sized microrobots. In *2008 IEEE/RSJ Int. Conf. Intel. Rob. Sys.* 4030–4035 (IEEE, 2008).
21. Haldane, D. W., Plecnik, M. M., Yim, J. K. & Fearing, R. S. Robotic vertical jumping agility via series-elastic power modulation. *Sci. Rob.* **1**, eaag2048 (2016).
22. Hu, W., Lum, G. Z., Mastrangeli, M. & Sitti, M. Small-scale soft-bodied robot with multimodal locomotion. *Nature* **554**, 81–85 (2018).
23. Kim, Y., Yuk, H., Zhao, R., Chester, S. A. & Zhao, X. Printing ferromagnetic domains for untethered fast-transforming soft materials. *Nature* **558**, 274–279 (2018).
24. Jäger, P. *Cebrennus* Simon, 1880 (Araneae: Sparassidae): a revisionary up-date with the description of four new species and an updated identification key for all species. *Zootaxa* **3790**, 319–356 (2014).
25. Zhakypov, Z., Huang, J. L. & Paik, J. A novel torsional shape-memory alloy actuator: modeling, characterization, and control. *IEEE Robot. Autom. Mag.* **23**, 65–74 (2016).
26. Wood, R.J, Avadhanula, S., Sahai, R., Steltz, E. & Fearing R. S. Microrobot design using fiber-reinforced composites. *J. Mech. Des.* **130**, 052304 (2008).
27. Zhakypov, Z. & Paik J. Design methodology for constructing multi-material origami robots and machines. *IEEE Trans. Rob.* **34**, 151–165 (2018).
28. Felton, S., Tolley, M. Demaine, E., Rus, D. & Wood, R. J. A method for building self-folding machines. *Science* **345**, 644–646 (2014).
29. Bennet-Clark, H. C. The energetics of the jump of the locust *Schistocerca gregaria*. *J. Exp. Biol.* **63**, 53–83 (1975).
30. Nabawy, M. R., Sivalingam, G., Garwood, R. J., Crowther, W. J. & Sellers, W. I. Energy and time optimal trajectories in exploratory jumps of the spider *Phidippus regius*. *Sci. Rep.* **8**, 7142 (2018).

Acknowledgements We thank H. Aonuma from the Complex Systems Research Group of Hokkaido University for providing insight into the behaviours of the trap-jaw ant. K.M. acknowledges financial support from the Japan Public–Private Partnership Student Study Abroad Program. This work is supported by the Swiss National Science Foundation (SNSF) 'START' Project and the Swiss National Center of Competence in Research (NCCR) Robotics.

Reviewer information *Nature* thanks Adam Stokes and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions Z.Z., K.M., K.H. and J.P. designed the study and interpreted the results. Z.Z. and J.P. conceived the idea of developing multi-locomotion millirobot collectives. Z.Z. and K.M. developed the millirobot hardware platform. Z.Z. designed the multi-locomotion mechanisms and models, and K.M. designed the electronics and communication. Z.Z. produced the figures and videos. Z.Z. and J.P. wrote the manuscript with input from all authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1388-8.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1388-8.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to J.P.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Design of robot locomotion mechanisms. The snap-through and bending properties of the central flexible Y-hinge and the appropriate arrangement of the robot legs and latches by selective activation of the SMA actuators generate the height, distance and somersault jumping, walking or crawling gaits (Fig. 1d). These can be programmed on the onboard microcontroller or controlled remotely via a keyboard on a portable computer by setting the actuator activation sequence, duration and power through a custom graphical user interface. For height, distance and somersault jumping and for walking, the activation pattern (shown in Fig. 1d by the red-highlighted springs) transits the robot from an initial rest state to its stance and then to its take-off phases, but there are no presets for the flight and landing phases. The robot follows a ballistic projectile motion for all four manoeuvres after take-off, with different launch angle, velocity and body rotation during flight. We model the Y-hinge as a pin joint connecting three independently rotating legs (Extended Data Fig. 2). For distance jumping, somersault jumping and walking, the snap-through motion at the Y-hinge side closes two side legs, pushing the third rear leg against the ground at an angle α, and the ground reaction force lifts the robot in the air. For height jumping, the snap-through of the Y-hinge bottom rapidly closes the two bottom legs against the ground, which in turn produce a ground reaction force that launches the robot vertically upwards. The kinematics of the robot flight between take-off and landing (first touchdown) in x–y Cartesian coordinates are governed by a ballistic projectile motion given by

$$y = y_0 + x\tan\alpha - \frac{gx^2}{2v_0^2\cos^2\alpha} \tag{1}$$

where y_0 is the initial height of the robot, measured between the centre of the Y-hinge and the ground at the instant of take-off, α and v_0 are the launch angle and velocity, respectively, and g is the gravitational acceleration. Because of the high uncertainty in predicting the surface area of the robot for each gait, especially owing to free-body rotation during somersault jumping and walking, we omit air-drag effects. The maximum horizontal travel distance, d, and vertical height, h, are then calculated by

$$h = \frac{v_0^2\sin^2\alpha}{2g},\ d = \frac{v_0^2\sin 2\alpha}{g} \tag{2}$$

The jumping displacement of the robot is maximized by increasing v_0 and attaining $\alpha = \pi/2$ for height jumps and $\alpha = \pi/4$ for distance and somersault jumping and walking. Although Tribot does not rotate considerably when performing height and distance jumps (at least during ascent), it performs, on average, a two-thirds body rotatation for somersault jumping and a one-third body rotation for walking.

For crawling locomotion, the robot is in contact with the ground on its latched edges; the activation sequence using the linear and torsional actuators produces a periodic stick-slip motion (Fig. 1d, Extended Data Fig. 2c). In this sequence, the SMA torsional actuators raise the rubber pads to slip and drop them to stick, varying the contact friction. The crawling step is then calculated by

$$C = 4l(\sin\theta_o - \sin\theta_c) \tag{3}$$

where C is the crawling step size for a single stick-slip manoeuvre, l is the half-length of the leg, and θ_o and θ_c are the opening and closing angles between the bottom two legs (2 and 3), respectively, with $\theta_o = \theta_c + \theta_2$ and $\theta_2 = \theta_3$.

To establish a generalized robot multi-locomotion dynamic model and to determine its velocity for each gait, we adopt a Euler–Lagrange method. Employing a Newtonian ($F = ma$) approach is difficult for modelling not only multi-locomotion and multi-degree-of-freedom mechanisms, but also single-locomotion mechanisms. The energy-based approach of the Euler–Lagrange method provides insight into the locomotion mechanism performance in terms of stored energy and produced motion, and therefore enables design optimization of the system components that are responsible for motion. The total kinetic (E_K) and potential (E_P) energies of the robot are given by

$$E_K = \frac{1}{2}\sum_{i=1}^{3} m_i v_i^2 + \frac{1}{2}\sum_{i=1}^{3} J_i\omega_i^2 \tag{4}$$

$$E_P = \sum_{i=1}^{3} m_i g h_i + \frac{1}{2}\sum_{i=1}^{3} k_i s_i^2 \tag{5}$$

where m_i, J_i, v_i, ω_i and h_i are the mass, moment of inertia, Cartesian and angular velocities and height of the ith link (leg), respectively (Extended Data Fig. 2). k_i and s_i are the stiffness coefficient and the deflection of the ith SMA spring actuator, respectively. The Lagrangian function is then $L = E_K - E_P$. Because the high-speed (snap-through) rotation of links 1 and 2 and the low-speed (bending) rotation of links 2 and 3 produce all five gaits, the equations of motion, $\ddot{\theta} = \begin{bmatrix} \ddot{\theta}_i \\ \ddot{\theta}_{i+1} \end{bmatrix}$, where $i = 1$ for the jumping and walking gaits and $i = 2$ for crawling, can be computed by solving

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{\theta}}\right) - \frac{\partial \mathcal{L}}{\partial \theta} = 0 \tag{6}$$

The masses and moments of inertia in equations (4), (5) for all three links are equal and constant, and so it is the spring actuator stiffness coefficients k_i that define the stored energy and thus determine the velocity of the links when they are released at the moment of snap-through. For our robot, the stiffness balance is set as $k_1 + k_2 > k_3$ followed by $k_3 > k_1 + k_2$ to generate the snap-through for the jumping and walking gaits, and $k_1 > k_2 + k_3$ followed by $k_2 + k_3 > k_1$ for crawling (Extended Data Fig. 2). The parameters of the actuators are tailored at design, but are also controllable during operation by changing the input Joule heating power, which varies the martensitic and austentitic temperature of the SMA, and hence also the stiffness[27] (see section 'Actuation design').

As the structure of Tribot necessitates a different configuration and orientation for each locomotion, we define a local coordinate frame (x'–y') and global coordinate frame (x–y), which are related through a 2×2 rotation matrix and the angle α (Extended Data Fig. 2). As the position of the robot's centre of mass varies substantially between the stance and take-off states, we fix the origins for both coordinate systems at the centre of the Y-hinge. Then, the positions ($\boldsymbol{p}_i$) translational velocities (v_i) and acceleration of links 1 and 2 are calculated by

$$\boldsymbol{p}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} = \begin{bmatrix} l(\sin(\theta_i + \alpha) - \sin\alpha) \\ -l(\cos(\theta_i + \alpha) - \cos\alpha) \end{bmatrix},\ i = 1, 2 \tag{7}$$

$$v_i = \sqrt{\dot{x}_i^2 + \dot{y}_i^2} = l\dot{\theta}_i,\ i = 1, 2 \tag{8}$$

$$a_i = l\ddot{\theta}_i,\ i = 1, 2 \tag{9}$$

where $\dot{\theta}_i$ and $\ddot{\theta}_i$ are equivalent to angular velocity ω_i and acceleration $\dot{\omega}_i$, respectively. For height and distance jumping, $\theta_1 \approx \theta_2$, and for somersault jumping and walking, $\theta_1 > \theta_2$, owing to the brief activation of the bottom spring actuator before the snap-through, which limits the angular rotation of link 2. The overall robot velocity is then a sum of individual link velocities

$$m_T v = \sum_{i=1}^{2} m_i v_i \tag{10}$$

Here, m_T and $v = v_0$ are the mass and velocity of Tribot. The total actuation power P required to accelerate Tribot to a distance Δp between the stance and take-off can be calculated using

$$P = \frac{E_K}{\Delta p} v = \frac{m_T}{2}\frac{v^3}{\Delta p} \tag{11}$$

where E_K is the robot's kinetic energy. We can also calculate the COT for each locomotion by

$$\mathrm{COT} = \frac{E_K}{m_T r} = \frac{v^2}{2r} \tag{12}$$

COT is measured in J kg^{-1} m^{-1} and r is the total travel distance between take-off and landing positions, corresponding to h for the height jump, to d for distance jumping, somersaulting and walking, and to the step size C for crawling locomotion. Tribot can jump a horizontal distance of 230 mm on average with a take-off velocity of 1.44 m s^{-1}, resulting in a low COT.

Actuation design. The previously mentioned locomotion mechanisms are produced by compressing and storing energy in the three SMA spring actuators that are Joule heated by passing a direct current, and in the two SMA torsional sheet actuators, motion is activated by the thermal conduction of the heat that is generated by the micro-heaters. To enable both fully automatic and remotely controlled activation of the actuators, with tunable power, we use a pulse-width-modulation method. We adjust the average electrical power P_{avg} to each actuator by controlling the switching duty cycle, $0 \le \text{duty} \le 1$, of five metal–oxide–semiconductor field-effect transistors in the software as $P_{avg} = \text{duty} \times P_s$. The supply power P_s is governed by Kirchhoff's rule so that $P_s = \frac{V_s^2}{R_{SMA}}$, where V_s is the battery supply voltage and R_{SMA} is the electrical resistance of either the SMA spring or the copper micro-heater layer of the SMA torsional sheet. In reality, the electrical resistance of the SMA slightly increases with temperature; however, for the plots in Fig. 3a, we fix it to 2.2 Ω, which is measured at room temperature (that is, when it is

martensite). Also, the SMA actuators stiffen proportionally to the applied heat between the martensitic (colder) and austenitic (hotter) temperatures. For the spring actuators, shear stress dominates owing to coil twisting, and its force-to-deflection relation for the martensitic and austenitic phases can be approximated by a linear relation without considering the detwinning effect

$$F = \frac{G(T)q^4}{8D^3N}\delta \tag{13}$$

Here, $G(T)q^4/(8D^3N) = k$ is the SMA spring stiffness coefficient, $\delta = \frac{\pi ND}{\cos\varphi_i}(\sin\varphi_f - \sin\varphi_i)$ is the spring deflection, q is the diameter of the wire, D is the coil diameter, N is the number of turns, φ_i and φ_f are the coil initial (compressed) and final (extended) pitch angles, and the shear modulus G is a function of the temperature T and varies between the martensitic minimum, G_M, and the austenitic maximum, G_A. The compression force increases substantially with increasing q; however, higher currents are then required to heat the wire owing to the reduced resistance. This is an important trade-off in designing the actuators to generate sufficient force at low power for untethered applications. To ensure that Tribot can operate without an external power supply, we designed the spring actuators with $q = 0.25$ mm, $D = 0.9$ mm, $N = 32$ or 33, $G_A \approx 18$ GPa and $G_M \approx 7$ GPa.

Unlike the linear spring actuator, the torsional actuator generates a bending moment; therefore, normal stress dominates its behaviour. Assuming pure bending of a thin sheet, the torque-to-angular-deflection relation can be approximated by

$$\tau = \frac{Y(T)I}{u}\theta \tag{14}$$

Here, τ is the torque, Y is the temperature-dependent elastic modulus, which is in the range $Y_M \le Y \le Y_A$ (between the martensitic and austenitic elastic moduli), $I = \frac{wt^3}{12}$ is the second moment of inertia of a rectangular sheet with cross section $w \times t$, u is the length of the curved section of the actuator and θ is the sheet bending angle. For our torsional actuators[24], $t = 0.1$ mm, $w = 8$ mm, $u = 6.5$ mm, $Y_A \approx 34$ GPa and $Y_M \approx 19$ GPa. The external heater layer has a resistance of 7 Ω, is thin (<0.05 mm) and consumes power as low as 0.5 W.

Experimental design. To measure the robot's height-, distance- and somersault-jumping trajectories and its walking and crawling steps, we set up eleven different experimental scenarios (Fig. 2, Extended Data Table 2). Each locomotion experiment is video-recorded and analysed using an open-source scientific video-tracking software called Tracker (https://physlets.org/tracker/). As the robot's leg snap-through motion occurs within 100 ms, we used a high-frame-rate camera (Sony Cyber-shot DSC-RX100 IV) with a recording speed of 250 fps to capture the robot's displacement in the x–y plane. The camera was configured to a real-time (25 fps) recording speed for the walking, crawling and multi-locomotion parkour experiments. We used a ruler (SI units) to calibrate the captured videos in the x or y axis, which were then analysed in Tracker. For all locomotion manoeuvres, we tracked the robot's central Y-hinge in x–y Cartesian coordinates from the instant of take-off to the instant of landing (first touchdown). We performed eleven independent experiments among all five gaits, each repeated six times. At the start of each run, Tribot was brought to an initial stance position. As each experiment measures twelve datasets (six each of x and y positions), with the data points not aligned in either of the axes, we interpolated each dataset using the Matlab *pchip* shape-preserving piecewise cubic interpolator function to align them in the x coordinate and then compute the mean for each experiment. We also calculated the average standard deviations in the y axis by taking the square root of the mean of variances (Extended Data Table 2). The standard deviation is plotted as a shaded region in each of the locomotion plots (Fig. 3a–c), using the Matlab *fill* function.

In the height-jump experiments (Figs. 2a, 3a), the robot is tested on two different edges: the sides with latches in contact with the ground and two sides with no latches (Supplementary Fig. 1b, Supplementary Video 1). The robot is controlled remotely for both experiments using a keyboard with a pre-set actuator activation power, displayed on a custom graphical user interface. For the loaded distance-jump experiment, we mounted a 5-g M8 stainless steel hexagonal nut at the robot's rear leg (Supplementary Video 1). To evaluate the robot's walking step size (Figs. 2d, 3b), we placed it into a 32-mm-wide channel made of transparent acrylic material to confine its lateral deviation while it flipped (Supplementary Video 2). For testing on a rough surface, we filled the channel floor with raisin-sized grains (FEPA F4 standard grain); they were removed for the smooth surface test. To evaluate the robot walking on a slope, we placed the channel on a smooth, inclined medium-density fibreboard with a slope of 10°.

To test the efficacy of the robot's crawling gait on different surfaces and on a slope, we programmed the robot to crawl with multiple steps on its edge with latches (Figs. 2e, 3c). We tested three terrain conditions: on sandpaper with roughness P100 (FEPA standard) and on medium-density fibreboard with a smooth finish, positioned horizontally (slope = 0°) and then inclined to 10°. Tribot crawls by periodically applying friction on the ground surface with rubber pads and sliding with the contact surface of the SMA torsional sheet actuator exposed after movement of the latch above the ground (a stick-slip movement), and so the surface interaction is essential in defining the crawling performance. We did not observe any horizontal propagation in the sandpaper test, owing to the increased friction between the torsional sheet actuator surface and the sandpaper, but the robot could crawl on the fibreboard with repeatable steps and even could crawl up a slope—although owing to sliding, it crawled with smaller steps on the slope than on a flat surface.

We computed the COT for each of Tribot's locomotion gaits (equation (12), Extended Data Table 3) and compared the distance-jumping COT to that of other multi-locomotion robots and insects, using take-off velocity, travel distance and mass data reported in the literature or extrapolated using equation (2). This comparison allowed effective benchmarking of engineered and biological systems in terms of locomotion efficiency and performance.

Robot fabrication. Robot hardware design for mass production should ideally be low-cost and customizable, for example using PCBs, which can be used to assemble diverse layouts of electronic components with versatile functionality in a matter of seconds. However, unlike PCBs, the mechanical design of several custom mechanisms and structural components dominates conventional robot construction, requiring meticulous assembly. Tribot's fabrication process allows robot multiplicity with minimal assembly effort (Fig. 1f). The robot's structure consists of two layers: a 300-μm-thick double-sided FR-4 PCB for structural backing and electronics and a 50-μm-thick Kapton polyimide film (DuPont) for the hinges, a material that is flexible and durable. The PCB workshop of the Swiss Federal Institute of Technology Lausanne mass-produced the PCB layer and the Kapton was cut on a laser micro-machining station (LAB 3550, Inno6). The two layers were bonded together using an adhesive film (Poli-Melt 701, Poli-Tape) and heat-pressed (Carver 3853CE, Carver) for 2 min at 160 °C with 90 N pressure. Then, to attach the electronic components, we applied a solder paste (SMD291AX, Chip Quick) onto the 100-μm-thick Kapton stencil placed on top of the PCB using a spatula, filling the component footprints. The stencil was gently removed, and the surface-mounted device components—including two infrared proximity sensors (VCNL4010, Vishay), two infrared transceivers (TFBS4711-TT1, Vishay) for communication, a microcontroller (ATTINY4313-MU, Atmel) and connectors and switches, among 50 other electronic components—were manually pick-placed onto the footprints. We then placed the PCB sheet for 3 min on a hot plate at 200 °C for solder reflow. Then, three 3.7 V, 40 mA h rechargeable lithium ion polymer batteries (DTP301120, Datapower) were soldered onto the terminals, and the two SMA torsional actuators with attached micro-heaters, plus four 3-mm rectangular pads moulded from silicone rubber (Elastocil M4601, Wacker Chemie AG), were glued onto the two latches. After cutting off the support bridges across the PCB hinge gaps (Fig. 1f), the multilayer sheet was folded to pop up into the robot's three-legged three-dimensional structure. Finally, we soldered a few wires to electrically connect one leg to the other, install the SMA spring actuators and test the assembled robot.

Our design uses off-the-shelf components and the total cost of each robot is under US$60. It takes approximately three hours for one skilled person to fabricate and manually assemble a robot. However, we could substantially reduce this time using an automated mass-production PCB assembly process. Our method facilitates processing of a wide range of materials with extremely fine features and greatly reduces the assembly effort, enabling low-cost and on-demand mass manufacturing of millirobots.

Communication range. Communication sensors allow multiple Tribot units to exchange information, interact and cooperate to execute collective tasks. Determining the sensor range helps to define the orientation and position of the next unit for a sustainable two-way (transmit–receive) communication link. The two infrared transceivers placed on either side of the robot's upper leg (the leg with no latch) produce a communication range of two symmetric sectors with a 60° angular opening, up to the maximum range of 1 m. For two robots to establish a two-way communication link, the maximum separation between them should not exceed 1 m and they should both be within the sector with orientation that meets the conditions

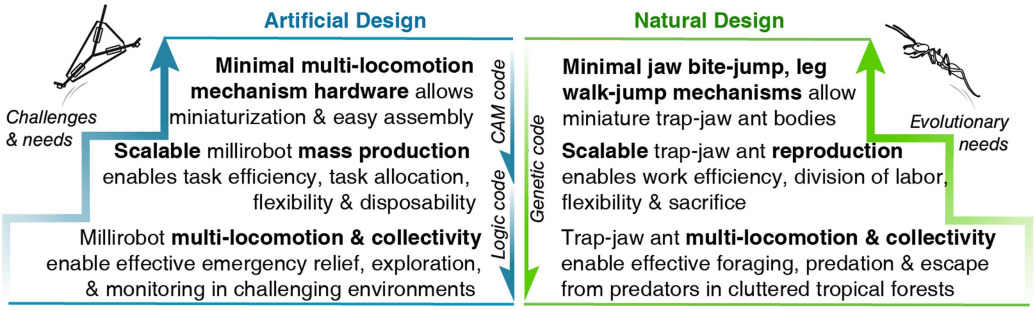
$$\beta_1 - \frac{\pi}{6} + n\pi \le \gamma \le \beta_1 + \frac{\pi}{6} + n,\ n = \ldots -1, 0, 1 \ldots \tag{15}$$

$$\gamma - \frac{\pi}{6} + m\pi \le \beta_2 \le \gamma + \frac{\pi}{6} + m\pi,\ m = \ldots -1, 0, 1 \ldots \tag{16}$$

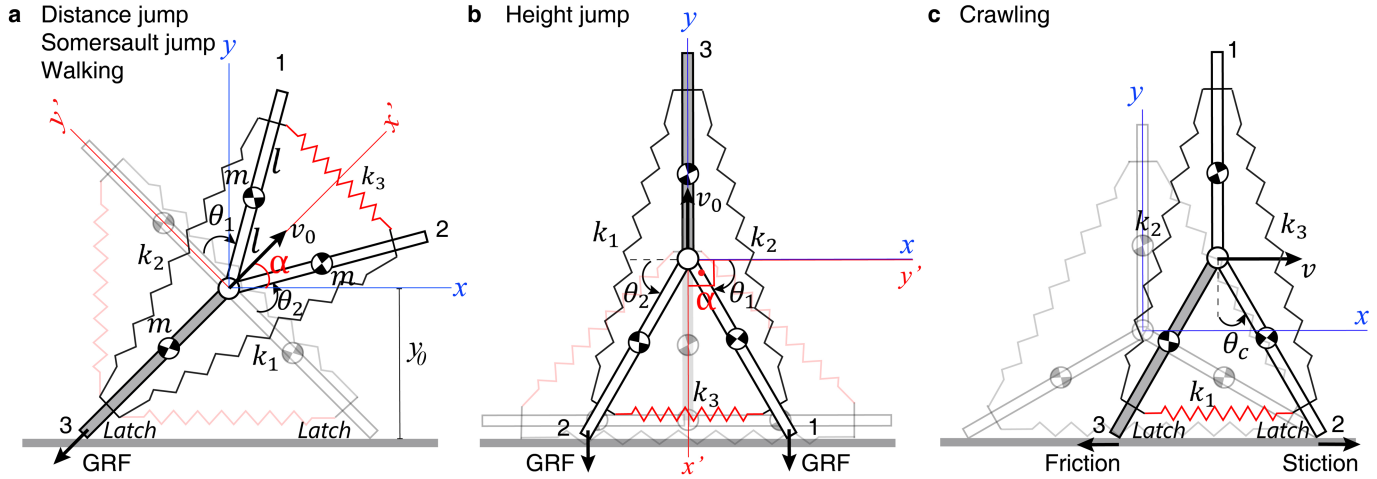
where β_1 and β_2 are the orientations of the first and the second robots on the ground x–z Cartesian coordinate plane, respectively, and γ is the relative angle between the two robots.

Data availability

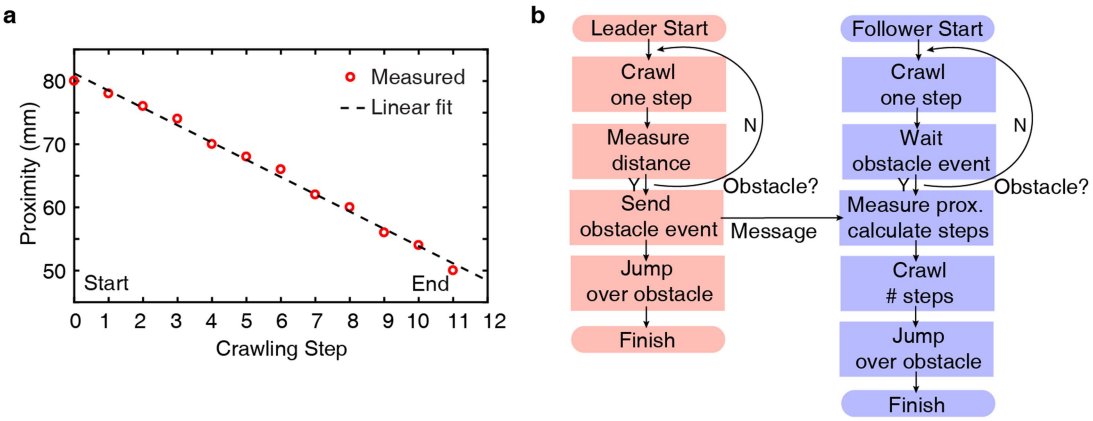
All data generated or analysed during this study are included in the published article, and are available from the corresponding author on reasonable request.



Extended Data Fig. 1 | Design challenges and needs of biological versus artificial multi-locomotion collectives. Trap-jaw ant collectivity, backed by scalable reproduction and the minimal multi-locomotion mechanisms that are integrated into their jaws and legs are the key to their survival in cluttered environments, which have emerged from evolutionary processes. Replicating these abilities in engineered systems will enable the use of millirobots in applications such as emergency mitigation, environmental monitoring and exploration with high task flexibility and efficiency. However, constructing minimal, integrated multi-locomotion mechanisms remains a major challenge for robotic hardware design that, when addressed, will enable robot miniaturization and assembly-free mass production for collective implementations. CAM, computer-aided manufacturing.



Extended Data Fig. 2 | Free-body diagrams of Tribot for calculating the locomotion kinematics and dynamics for all five locomotion gaits. **a**, Tribot transitioning from initial stance to take-off, applicable for the distance- and somersault-jump gaits and walking gaits as a result of snap-through at the Y-hinge side. The Y-hinge is modelled as a revolute pin joint connecting three links. The snap-through motion generated by compression of the SMA spring actuator (k_3) instantly rotates the side legs (1 and 2) and pushes the rear leg (3) against the ground. This produces a ground reaction force that lifts the robot in a ballistic projectile motion with take-off velocity v_0. For somersault jumping, the bottom spring actuator (k_1) activates shortly before the side spring (k_3) to achieve free-body rotation during flight. **b**, The robot can perform height jumps on any three edges; however, for reaching high altitudes, it is most effective on the edge without rubber friction pads (which are located on legs 2 and 3). Here, the snap-through occurs at the Y-hinge bottom, and the rapid closing of the bottom legs produces a ground reaction force that lifts the robot up vertically. **c**, The crawling locomotion occurs on the edge with latches; the robot moves by opening and closing the bottom legs (2 and 3) using stick-slip motion. GRF, ground reaction force.



Extended Data Fig. 3 | The proximity measurement data for the division-of-labour experiment and the event chart for the tandem-running experiment. a, The proximity data measured by the monitor robot, showing the linear propagation of the workers with each pushing step. The object is moved its set distance, 30 mm. **b**, The event chart for the leader–follower tandem-running experiment with obstacle avoidance by communication.

Extended Data Table 1 | Tribot's functional components and mass budget

Part/Layer	Material	Quantity	Mass (g)
Rigid legs	Fiberglass-Copper	1	3.1
Compliant hinge	Polyimide	1	0.49
SMA Actuators	Nitinol	5	0.35
Latch friction pads	Elastocil	4	0.5
Electronics & wires	Silicon, copper	>50	1.56
Batteries	Lithium-Polymer	3	3.7
Robot Total			**9.7**

Extended Data Table 2 | Tribot's locomotion performance under different conditions

	Experimental Condition	Mean Displacement (mm)	Standard Deviation (%)
1	Height jump, w/o latch	140.6	5
2	Height jump, w/ latch	72.5	7.5
3	Distance jump *P*=3.7 W	230	7.2
4	Distance jump *P*=2.7 W	140	8.8
5	Distance jump *P*=3.7 W, *L*=5 g	110	21.1
6	Somersault jump *P*=3.7 W	h: 88, d: 100	12.5
7	Walking on flat surface *P*=1.7 W	48.8	10.8
8	Walking on rough surface *P*=1.7 W	31.1	12.9
9	Walking onto slope 10°, *P*=2.2 W	28.8	25.7
10	Crawling on flat surface	4.9	8
11	Crawling onto slope 10°	2.6	28.8

h, height; d, distance.

Extended Data Table 3 | Comparison of Tribot with small-scale terrestrial multi-locomotion robots and insects

Robot/Insect Name	Mass (g)	Size (mm)	Number of Locomotion	Energy Storage, Transmission & Actuation	Jump h./ Size	Jump d./ Size	COT ($J.kg^{-1}m^{-1}$)
JumpRoach [13]	59.4	100	2, crawling & jumping	6 rubber bands, 1 four-bar linkage leg, 3 gears, 3 DC motors	11.0	7.0	h: 9.7 d: 22.4
MSU Jump-Runner [14]	25	90	2, running & jumping	2 elastic strips, 1 slider-crank-cable leg, 8 gears, 2 DC motors	16.0	6.6	h: 9.8 d: 23.6
MiniWhegs [15]	146	90	2, running & jumping	1 steel spring, 2 four bar linkage legs, 2 chains, 2 DC motors, 4 Whegs	2.4	1.2	h: 10.5 d: 19.3
Surveillance Robot [16]	250	140	2, rolling & jumping	1 steel spring, 1 six-bar linkage leg, 4 gears, 3 DC motors, 2 wheels	2.9	2.9	h: 11.7 d:11.7
Wheel-Hopper [17]	170	168	2, rolling & jumping	1 steel spring, 2 four-bar linkage legs, 4 gears, 3 DC motors, 2 wheels	0.7	0.3	h: 10.7 d: 23.7
Trap-Jaw Ant *Odontomachus bauri [5]*	*0.01*	*13*	*3, jaw-jump, leg-jump, leg-walk*	*Flexor-extensor jaw snap catapult & leg muscles*	*5.6 escape 2.4 defense*	*2.4 escape 17.2 defense*	*h: 0.4 d: 0.9 escape*
Desert Locust *Schistocerca gregaria [29]*	*3*	*60*	*3, leg-jump, leg-walk, flying*	*Flexor-extensor leg muscle contruction*	9.8	*10.0*	*h: 6.2 d: 6.1*
Jump. Spider *Phidippus princeps [30]*	*0.15*	*15*	*2, leg-jump, leg-walk*	*Flexor-extensor leg muscle contruction*	0.3	*4.0*	*h: 69.8 d: 5.7*
Tribot (in this article)	**9.7**	**58**	**5, height jump, distance jump, somersault jump, walking & crawling**	**3 SMA springs & 2 SMA sheet, 3 tree linkage legs, 1 snap-bend Y-hinge, 2 rubber latches**	**2.4**	**4.0**	**h: 14.4 d: 4.5 s: 14.8 w: 4.4 c: <1**

Data are from refs [13–17]. DC, direct current; h, height; d, distance; s, somersault; w, walking; c, crawling. Wheg, wing-leg or wheel-leg.

LETTER

https://doi.org/10.1038/s41586-019-1331-z

Living annulative π-extension polymerization for graphene nanoribbon synthesis

Yuuta Yano[1], Nobuhiko Mitoma[1,2], Kaho Matsushima[1], Feijiu Wang[1,2], Keisuke Matsui[1,2], Akira Takakura[1,2], Yuhei Miyauchi[1,2,3], Hideto Ito[1,2]* & Kenichiro Itami[1,2,4]*

The properties of graphene nanoribbons (GNRs)[1–5]—such as conductivity or semiconductivity, charge mobility and on/off ratio—depend greatly on their width, length and edge structure. Existing bottom–up methods used to synthesize GNRs cannot achieve control over all three of these parameters simultaneously, and length control is particularly challenging because of the nature of step-growth polymerization[6–18]. Here we describe a living annulative π-extension (APEX)[19] polymerization technique that enables rapid and modular synthesis of GNRs, as well as control over their width, edge structure and length. In the presence of palladium/silver salts, *o*-chloranil and an initiator (phenanthrene or diphenylacetylene), the benzonaphthosilole monomer polymerizes in an annulative manner to furnish fjord-type GNRs. The length of these GNRs can be controlled by simply changing the initiator-to-monomer ratio, achieving the synthesis of GNR block copolymers. This method represents a type of direct C–H arylation polymerization[20] and ladder polymerization[21], activating two C–H bonds of polycyclic aromatic hydrocarbons and constructing one fused aromatic ring per chain propagation step.

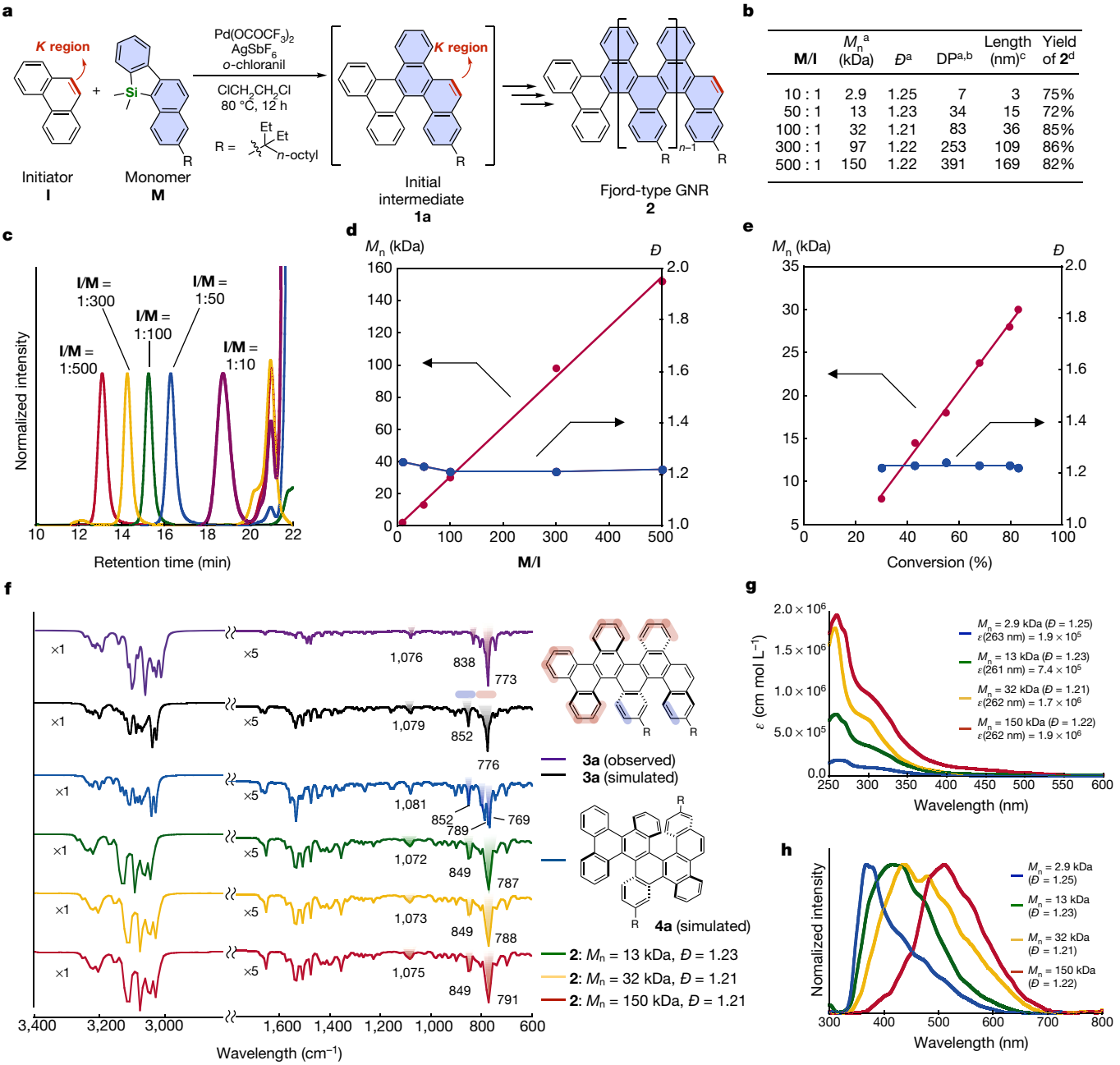
To achieve full synthetic control over the structures of GNRs, we must draw inspiration from organic synthesis, where a target molecular entity is built up with atom-by-atom precision[7–18]. Müllen[10–14], Fasel[11,13,14] and others[11,13–18] have reported bottom–up synthetic approaches, such as the solution-phase Suzuki–Miyaura coupling polymerization[10,18], the Diels–Alder polymerization[12] and on-surface polymerization[11,13–15,17] of small aromatic components to form polyarylene precursors, which are stitched into GNRs by cyclodehydrogenation (Scholl reaction). These methods have received considerable attention owing to their high potential of controlling the width (~1 nm) and edge structures of GNRs, especially with respect to top–down synthesis by lithographic cutting of graphene or unzipping of carbon nanotubes[6–9]. Whereas the width and edge structures can be controlled precisely in solution-phase syntheses, precise length control of GNRs is yet to be achieved. The same is true for on-surface GNR synthesis, where length control remains an unresolved issue[11,13–15].

Recently, we established APEX methodologies[19] for the rapid synthesis of nanographenes from simple aromatic templates[22,23]. For example, unfunctionalized polycyclic aromatic hydrocarbons can be directly coupled at the *K* regions (convex armchair edge) with dibenzosilole derivatives in the presence of a palladium catalyst[22,23]. We envisaged that the use of phenanthrene (**I**) as an initiator and benzonaphthosilole (**M**) as a monomer would result in a living[24] and controllable APEX polymerization (Fig. 1a). A crucial feature of this approach is the use of **M** as a monomer substrate. **M** contains a masked *K* region[22], which is only exposed upon carrying out the Pd-mediated APEX reaction. Thus, the initial APEX reaction occurs selectively between the *K* region of **I** and **M** to provide tribenzochrysene (**1a**) as the initial intermediate. This results in the formation of a new fused aromatic ring and unmasks a new *K* region, primed for subsequent APEX reactions. Because of the steric congestion around the *K* region of **1a** at the propagation terminus, it is expected that the reaction rate of propagation between **1a** and **M** is slower than the initiation reaction. Propagation would then continue in a regioselective manner to afford fjord-type GNR **2**.

After extensive optimization studies, we established the conditions for the living APEX polymerization. The benzonaphthosilole monomer **M** polymerized smoothly in the presence of **I**, $Pd(OCOCF_3)_2$, $AgSbF_6$ and *o*-chloranil in 1,2-dichloroethane at 80 °C to afford fjord-type GNR **2** (Fig. 1a). Analysis of the crude product by size-exclusion chromatography (SEC) with polystyrene (PS) standards indicated that **2** was monomodal with narrow dispersity under various **M**/**I** ratios (Fig. 1b, c). The estimated number-average molecular weight (M_n), dispersity ($Đ = M_w/M_n$; where M_w is weight-average molecular weight), degree of polymerization (DP) and estimated polymer length of the crude GNRs, as well as the yields of purified GNRs, are shown in Fig. 1b. This APEX polymerization enabled the length-controlled synthesis of fjord-type GNR **2** merely by rational changes to the **M**/**I** ratio. In the reaction with **M**/**I** = 10, the relatively short GNR **2** ($M_n = 2.9 \times 10^3$ Da, DP = 7) was obtained in a narrow range of *Đ* values. Increasing the **M**/**I** ratio resulted in the formation of a longer GNR **2**: the reaction with a **M**/**I** ratio of 500 furnished **2** in 82% yield with $M_n = 1.5 \times 10^5$ Da, $Đ = 1.22$ and DP = 391. We also conducted matrix-assisted laser desorption/ionization time-of-flight (MALDI–TOF) mass analysis of a representative sample (GNR **2** with $M_n = 1.3 \times 10^4$ Da and $Đ = 1.23$) to confirm the validity of the PS-based M_n, as conventional PS standards may be conformationally too flexible to reflect the real molecular weight of rigid GNRs[12]. MALDI–TOF mass analysis revealed eleven distinct mass peaks, each observed with gaps of $m/z = 382$ (*m*, mass; *z*, charge; Extended Data Fig. 1a, b), supporting the specific formation of the fjord-type GNR structure without any defects such as non-cyclized or non-aromatized products. Furthermore, the dispersity of the polymer ranged from DP(*n*) = 21–31 (m/z = 8,000–12,000), implying a narrow dispersity range in the current APEX polymerization. The m/z value of the peak top (m/z = 10,116.9099) was considered to reflect the real M_n of GNR **2**, which is slightly smaller than the PS-based M_n. Therefore, we used the PS-based M_n in the current study to estimate the molecular weight of GNRs.

The APEX polymerization was found to be highly regulated. The M_n value changed predictably with the **M**/**I** ratio, whereas *Đ* was virtually constant (~1.23) and independent of the **M**/**I** ratio (Fig. 1d). The M_n value was also completely proportional to conversions of **M** without any change of *Đ* (Fig. 1e). In the reaction with **M**/**I** = 10, the chain propagation stopped within 1 h, but the propagation resumed to produce the longer GNR **2** ($M_n = 3.3 \times 10^4$ Da) when additional **M** (90 equivalents relative to **I**) was added (Extended Data Fig. 1c, d). Notably, the M_n and *Đ* values of **2** were almost identical to those obtained in the reaction with **M**/**I** = 100 (Fig. 1b). Thus, the present APEX polymerization satisfies the requirements for living polymerization. Although there have been some reports of GNRs with low *Đ* values (1.06–1.12) after purification and fractionation by SEC[25], this work demonstrates the first living polymerization for GNRs with low *Đ* even in the crude mixture.

[1]Graduate School of Science, Nagoya University, Nagoya, Japan. [2]JST-ERATO, Itami Molecular Nanocarbon Project, Nagoya, Japan. [3]Institute of Advanced Energy, Kyoto University, Kyoto, Japan. [4]Institute of Transformative Bio-Molecules (WPI-ITbM), Nagoya University, Nagoya, Japan. *e-mail: ito.hideto@g.mbox.nagoya-u.ac.jp; itami@chem.nagoya-u.ac.jp



M/I	M_n[a] (kDa)	Đ[a]	DP[a,b]	Length (nm)[c]	Yield of 2[d]
10 : 1	2.9	1.25	7	3	75%
50 : 1	13	1.23	34	15	72%
100 : 1	32	1.21	83	36	85%
300 : 1	97	1.22	253	109	86%
500 : 1	150	1.22	391	169	82%

Fig. 1 | APEX polymerization for structurally well defined GNRs. **a**, Living APEX polymerization using phenanthrene (**I**) as the initiator and benzonaphthosilole (**M**) as the monomer. Reaction conditions: **M** (1.0 equiv.), $Pd(OCOCF_3)_2$ (1.0 equiv.), $AgSbF_6$ (2.0 equiv.), *o*-chloranil (2.0 equiv.), $ClCH_2CH_2Cl$, 80 °C, 12 h. **b**, Table of results of polymerization. **I**/**M**, molar ratio of **I** and **M**. [a]Determined by using PS standards. [b]Calculated using M_n. [c]Estimated length using the DP value. Analyses of M_n and Đ were conducted with the crude product of GNRs. [d]Isolated yield after purification by SEC. **c**, SEC charts of the crude products of reactions with various **M**/**I** ratios. **d**, **e**, M_n and M_w ($Đ = M_w/M_n$) profiles versus **M**/**I** and conversion of the monomer in the living APEX polymerization. **f**, IR spectra of dimer **3a** (black line) and possible isomer **4a** (blue line) predicted by density functional theory (DFT) calculations at the B3LYP/6-31G(d) level, and observed IR spectra of dimer **3a** (purple line) and GNR **2** with different polymer lengths (green, yellow and red lines). The IR spectra of **3a** and various sizes of GNR **2** support the formation of a fjord-type GNR structure (see section 12 in Supplementary Information for details). **g**, Absorption spectra of GNR **2** with different polymer lengths. The molar absorptivity ε was determined using the M_n values measured by SEC analyses. **h**, Fluorescence spectra of GNR **2** with different polymer lengths.

The structure of fjord-type GNR **2** was carefully analysed by infrared (IR), Raman, nuclear magnetic resonance (NMR), ultraviolet–visible (UV-vis) absorption and emission spectroscopies to confirm the high degree of regiocontrol during polymerization (Fig. 1f–h, section 12 in Supplementary Information; see Extended Data Figs. 2a, 3 for details of the Raman and NMR spectroscopic analyses). After the initiation reaction produces the first intermediate (**1a** in Fig. 1a), the subsequent APEX reaction can in principle provide two possible regioisomeric dimers (**3a** and **4a** in Fig. 1f). Thus, we conducted step-by-step APEX reactions to synthesize dimer **3a** and model dimer **3b** (Extended Data Fig. 2b, c), which are more plausible products having [5]helicene moieties, for reliable spectral comparisons. The stoichiometric APEX reaction of **I** with model monomer **M′** afforded tribenzochrysene (**1b**) in 91% yield, which is further π-extended with **M′** to provide **3b** in 86% yield. In this reaction, the structural isomer **4b**, which has sterically congested [6]helicene moieties (which typically show upfield-shifted hydrogen atoms on [6]helicene regions at 6–7 p.p.m. in ^{1}H NMR) was not obtained at all (see ^{1}H NMR spectrum of **3b** in Supplementary

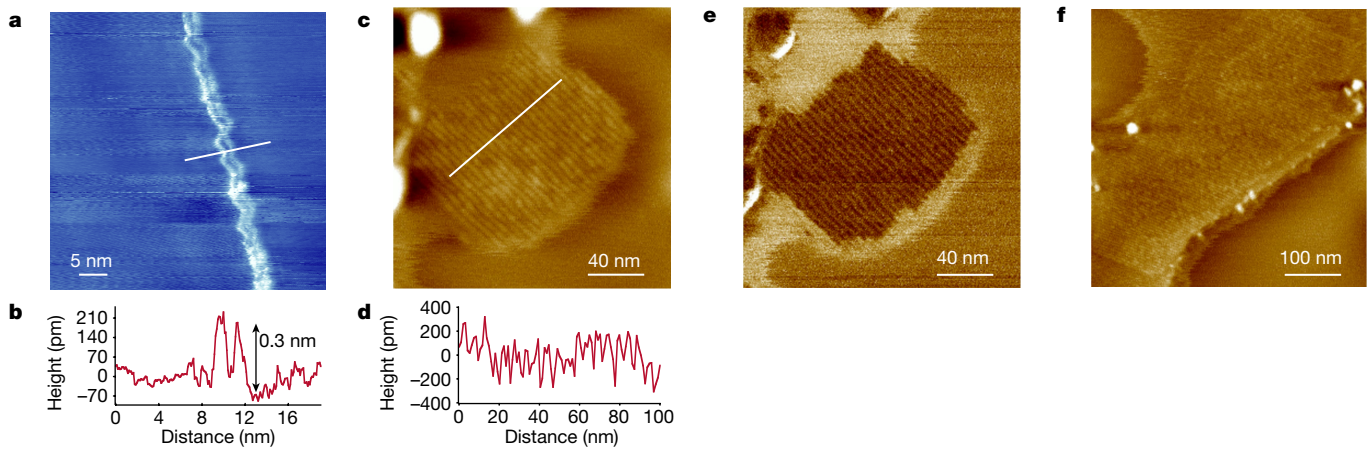
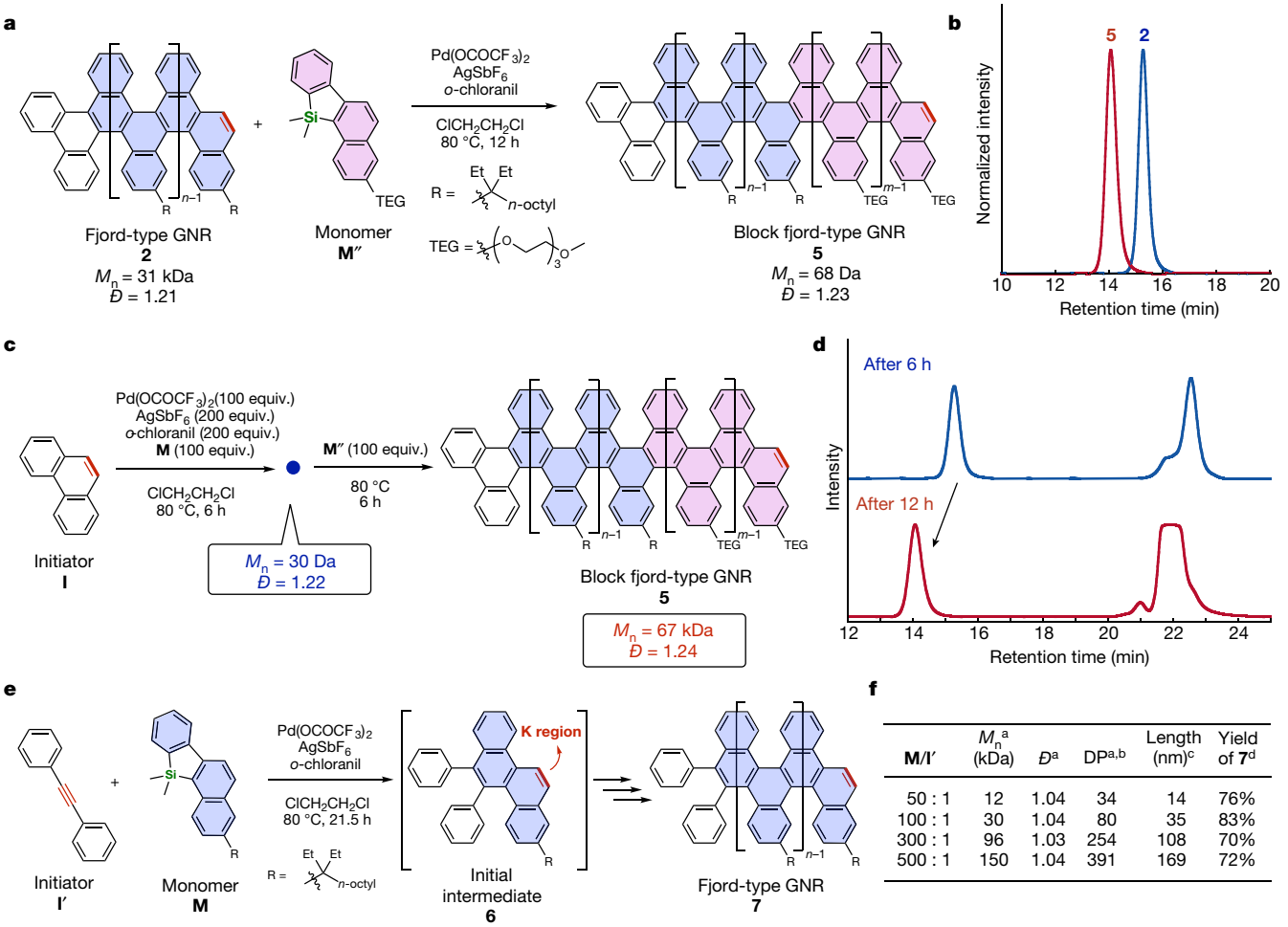
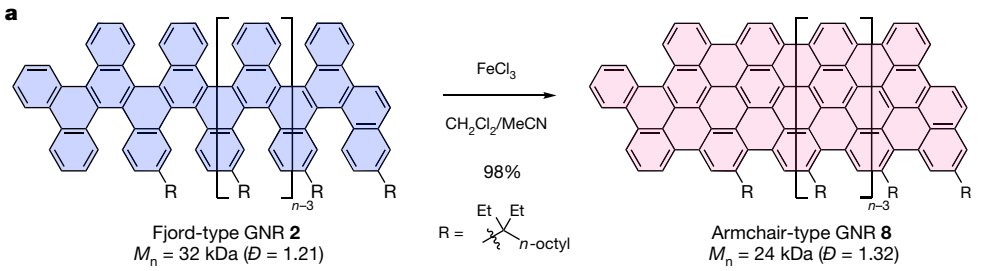


Fig. 2 | STM and AFM images of fjord-type GNR 2. a, STM image of fjord-type GNR **2** ($M_n = 1.5 \times 10^5$ Da, $Đ = 1.22$) deposited on HOPG (voltage $U = 1.0$ V, current $I = 50$ pA, temperature $T = 78$ K). **b**, Cross-sectional height profile taken along the white line in **a**, perpendicularly to the observed stripes. **c**, AFM height image of GNR **2** ($M_n = 1.5 \times 10^5$ Da, $Đ = 1.22$) on HOPG. **d**, Cross-sectional height profile taken along the white line in **c**, perpendicularly to the observed stripes. **e**, AFM phase image of **c**. **f**, Height image of submicrometre-scale self-assembly of GNR **2** ($M_n = 1.5 \times 10^5$ Da, $Đ = 1.22$).



M/I′	M_n[a] (kDa)	$Đ$[a]	DP[a,b]	Length (nm)[c]	Yield of 7[d]
50 : 1	12	1.04	34	14	76%
100 : 1	30	1.04	80	35	83%
300 : 1	96	1.03	254	108	70%
500 : 1	150	1.04	391	169	72%

Fig. 3 | Living APEX block copolymerization and highly controlled living APEX polymerization. a, Living APEX block copolymerization using GNR **2** as the initiator and benzonaphthosilole **M″** as the monomer. Reaction conditions: **2** (1.0 equiv.), **M″** (100 equiv.), $Pd(OCOCF_3)_2$ (100 equiv.), $AgSbF_6$ (200 equiv.), *o*-chloranil (200 equiv.), $ClCH_2CH_2Cl$, 80 °C, 12 h. **b**, SEC charts before and after the reactions. **c**, Confirmation of livingness of APEX polymerization of **I**, **M** and **M″** by one-pot SEC analyses. **d**, SEC charts of the reaction progress after 6 and 12 h. **e**, Living APEX polymerization using diphenylacetylene (**I′**) as the initiator and **M** as the monomer. Reaction conditions: **M** (1.0 equiv.), $Pd(OCOCF_3)_2$ (1.0 equiv.), $AgSbF_6$ (2.0 equiv.), *o*-chloranil (2.0 equiv.), $ClCH_2CH_2Cl$, 80 °C, 21.5 h. **f**, Results of polymerization with various molar ratios of **M**/**I′**. [a]Determined using PS standards. [b]Calculated using M_n. [c]Estimated length using DP. All analyses of M_n and $Đ$ were conducted with the crude product of GNRs. [d]Isolated yield after purification by SEC.



b

GNR **2**		GNR **8**		
M_n[a] (kDa)	$Đ$[a]	M_n[a] (kDa)	$Đ$[a]	Yield (%)[b]
2.9	1.25	1.4	1.31	95
13	1.23	8.3	1.31	96
32	1.21	24	1.32	98
97	1.22	68	1.35	95
150	1.22	120	1.34	96

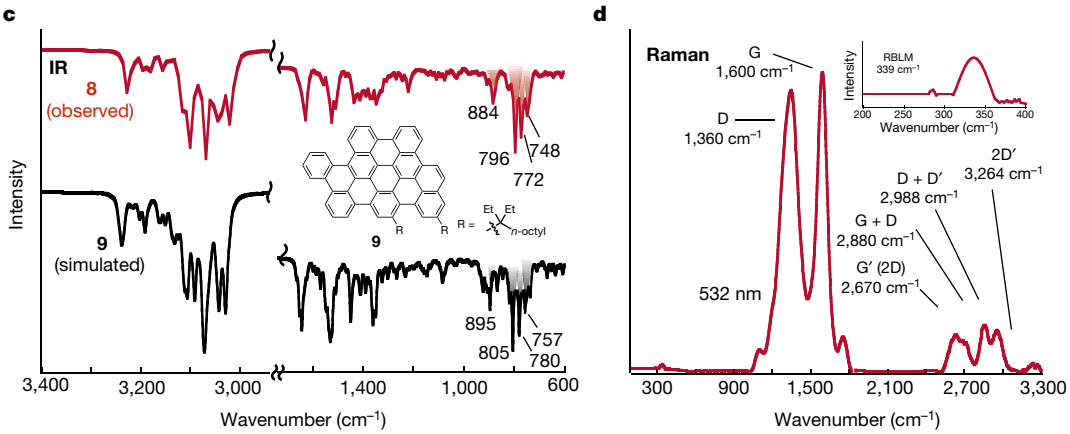


Fig. 4 | Transformation of fjord-type GNR 2 to armchair-type GNR 8. **a**, Scholl reaction of **2** in the presence of $FeCl_3$ (7.0 equiv. with respect to the number of hydrogen atoms in the fjord regions). **b**, Results of all Scholl reactions of GNR **2** with various polymer lengths. [a]Determined using PS standards. [b]Isolated yield after purification. **c**, Observed IR spectra of **8** (red line) and DFT-calculated (B3LYP/6-31G(d)) IR spectra of dimeric armchair GNR **9** (black line) (see section 12 in Supplementary Information for details). **d**, Raman spectrum of **8**, obtained with by a 532-nm excitation laser. In the Raman spectrum of **8**, a sharp and intense G-band peak and a larger D-band peak were observed at 1,600 and 1,360 cm^{-1}, respectively, which are typical for armchair-type GNRs[17]. Well resolved double-resonance signals were also observed at 2,670, 2,880, 2,988 and 3,264 cm^{-1}, which were assigned to G′ (2D), G + D, D + D′ and 2D′ peaks, respectively. Furthermore, a radial breathing-like mode (RBLM) stretch, in which the wavenumber is well known to correspond to the width of the armchair-type GNR[29,30], was observed at 339 cm^{-1}. This value was in between those observed in armchair-type GNRs with $N = 7$ (396 cm^{-1})[11] and 9 (312 cm^{-1})[14] and in good agreement with the calculated value of armchair-type GNRs with $N = 8$ (328 cm^{-1})[29].

Information). Dimer **3a** was also successfully synthesized in 67% yield in the same fashion (Extended Data Fig. 2c). Although the reason for this excellent regioselectivity is not fully understood at this stage, each regioselective APEX propagation step was observed to enable highly controlled polymerization.

The absorption and fluorescence spectra of all sizes of GNR **2** show that they were both red-shifted as GNRs became longer (Fig. 1g, h). In particular, wide-range emission covering the visible region (wavelength $\lambda = 400–700$ nm) was observed in longer GNRs with $M_n = 1.5 \times 10^5$ Da. Thus, we have revealed that a length effect exists in the photophysical properties of GNRs, especially in fluorescence.

The highly soluble nature of GNR **2** enables microscopic visualization, such as scanning tunnelling microscopy (STM) and atomic force microscopy (AFM) of individual molecules and self-assembled structures, respectively, by simply evaporating GNR solutions of 1,2,4-trichlorobenzene on highly oriented pyrolytic graphite (HOPG) (Fig. 2, Supplementary Figs. 1–11). After deposition of GNR **2** on HOPG, measurements by STM showed the existence of a pair of two waving wires (Fig. 2a, Supplementary Fig. 1). Judging from the height (~0.3 nm) and widths of the single- and double-waving wires (1.7–2.2 nm and 3.9 nm, respectively), these are considered to be two isolated fjord-type GNR **2** wires that appear to be flexible, arising from the helical and waggling conformations (Fig. 2b)[23]. In other STM measurements, we obtained various self-assembly images (Supplementary Figs. 1–11), such as rope-like helical wires of GNR **2** (Supplementary Figs. 6, 7) and a longitudinally assembling long straight wire of GNRs **2** (Supplementary Figs. 3, 4, 9–11). We also observed by AFM uniform stripe images of close-packed GNR **2** with intervals of about 5 nm (Fig. 2c–f). The longer wires might be bundled GNR **2**. Although the detailed assembly model remains unclear, such diversity in the assembly pattern indicates attractive applications in materials science and supramolecular chemistry.

The living nature of APEX polymerization allows us to synthesize a range of previously inaccessible molecular nanocarbon materials. For example, living APEX block copolymerization was also possible by reacting methoxytriethylene glycol (TEG)-substituted benzonaphthosilole monomer **M″** to fjord-type GNR **2** with $M_n = 3.1 \times 10^4$ Da ($Đ = 1.21$) (Fig. 3a, Extended Data Fig. 4). The polymerization afforded A_n–B_m type block copolymer **5** with $M_n = 6.8 \times 10^4$ Da ($Đ = 1.23$). The introduction of a TEG group on the monomer did not affect the high efficiency and narrow dispersity range achieved in the synthesis of **2** with a 3-ethyl-3-undecyl group (Fig. 3b). The one-pot synthesis of the GNR block copolymer was also possible, starting from a phenanthrene initiator **I** with a monomer **M**, followed by sequential addition of a second monomer **M″** (Fig. 3c, d). The M_n and $Đ$ values of GNRs before and after the addition of monomer **M″** were almost identical to those obtained in the experiments shown in Fig. 3a. These results are promising for the prospects of living APEX polymerization, such as application to highly precise synthesis of graft and star copolymers, fine-tuning of polymer units and properties, and supramolecular assembly using amphiphilic GNR polymers.

Although the proposed APEX polymerization system certainly yields precise GNRs, there still exists room for improvement in terms of dispersity (length control). Unlike the well developed living polymerization of olefins[26], APEX polymerization provides GNR polymers with up to $Đ = 1.21$. Inspired by our previously developed Pd-catalysed APEX reaction of diarylacetylenes with dibenzosilole[27], we tested

diphenylacetylene (**I′**) as an alternative initiator for living APEX polymerization (Fig. 3e). We expected initiator **I′** to react with monomer **M** more rapidly than **I** to provide the diphenylchrysene intermediate **6**. To our delight, the living APEX polymerization of **I′** and **M** effectively proceeded to afford diphenyl-substituted fjord-type GNR **7** (Fig. 3e, Extended Data Fig. 5). Amazingly, the synthesis of low-dispersity GNR with $Đ = 1.03–1.04$ was accomplished with various ratios of **M**/**I′** (Fig. 3f). The MALDI–TOF mass measurement also supported these rather low $Đ$ values of GNR **7** (Extended Data Fig. 6). Such an extremely low $Đ$ of GNR has never been realized by other synthetic methods. The use of diarylacetylenes as an initiator would have great benefits, not only because of the small $Đ$, but also because of the structural diversity and the fine-tunability of the terminal structure of GNRs.

We also attempted to convert the length-defined fjord-type GNRs **2** to the thicker, armchair-type GNR **8**—the so-called $N = 8$ armchair GNR (where N represents the number of carbon atoms within the width of the GNR)[17]—by cyclodehydrogenation (Scholl reaction) with $FeCl_3$ (Fig. 4, Extended Data Figs. 7, 8). Through the Scholl reaction, the PS-based M_n ($Đ$) changed from 3.2×10^4 Da (1.21) to 2.4×10^4 Da (1.32) (Fig. 4b, Extended Data Fig. 8a). We also prepared $N = 8$ armchair-type GNR **5** with $M_n = 1.4 \times 10^3$ Da ($Đ = 1.31$), 8.3×10^3 Da (1.31), 6.8×10^4 Da (1.35) and 1.2×10^5 Da (1.34) from fjord-type GNR **2** with $M_n = 2.9 \times 10^3$ Da ($Đ = 1.25$), 1.3×10^4 Da (1.23), 9.7×10^4 Da (1.22) and 1.5×10^5 Da (1.22), respectively (Fig. 4b, Extended Data Fig. 7a). As in the case of GNR **2**, the formation of a thicker GNR structure in **8** was supported by STM and AFM measurements with interesting self-assembly patterns (see Extended Data Fig. 9, Supplementary Figs. 12–18 for details).

In the absorption spectra of **8** with various lengths (Extended Data Fig. 8b), the shoulder feature at longer wavelengths appeared from 400 nm to 1,200 nm, which implies the formation of a π-extended structure in **8**. The maximum emission wavelengths (λ_{max}) changed from 417 nm to 595 nm, and the broad emission peaks appeared in the visible and even in the near-infrared region, which clearly indicate the π-extension from **2** (Extended Data Fig. 8c). However, as the theoretically predicted pristine $N = 8$ armchair-type GNR has a tiny optical bandgap[28] of 0.42 eV, the data are suggestive of the presence of small defects (see Extended Data Fig. 7 for the evaluation of the efficiency of the Scholl reaction and defects). It is likely that trace defects in **8** (uncyclized fjord regions) caused the local bandgap opening, resulting in unexpected photoluminescence around visible-light regions. On the other hand, IR and Raman spectroscopic analyses of GNR **8** show changes in the peripheral structure after the transformation (see Fig. 4c, d, Extended Data Fig. 8d, e and section 12 in Supplementary Information for details). The inherent nature of GNR **8** and the effect of defects seems to be difficult to determine in the present work, so further investigation into improvements in the Scholl reaction, the synthesis of defect-free GNRs and measurements of conductivity/charge-carrier density will be conducted in the future.

The present study not only demonstrates the first precision synthesis of GNRs with simultaneous width, edge structure and length control, but also introduces a completely new type of polymerization, opening doors in polymer and materials science. The present modular GNR synthesis will enable researchers to explore the effect of length on the properties of GNRs and provide a range of tailor-made GNR-related molecular nanocarbon materials.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1331-z.

Received: 31 May 2018; Accepted: 1 May 2019;
Published online 26 June 2019.

1. Dutta, S. & Pati, S. K. Novel properties of graphene nanoribbons: a review. *J. Mater. Chem.* **20**, 8207–8223 (2010).
2. Li, X., Wang, X., Zhang, L., Lee, S. & Dai, H. Chemically derived, ultrasmooth graphene nanoribbon semiconductors. *Science* **319**, 1229–1232 (2008).
3. Son, Y. W., Cohen, M. L. & Louie, S. G. Half-metallic graphene nanoribbons. *Nature* **444**, 347–349 (2006); corrigendum 446, 342 (2007).
4. Yang, L., Park, C. H., Son, Y. W., Cohen, M. L. & Louie, S. G. Quasiparticle energies and band gaps in graphene nanoribbons. *Phys. Rev. Lett.* **99**, 186801 (2007).
5. Ritter, K. A. & Lyding, J. W. The influence of edge structure on the electronic properties of graphene quantum dots and nanoribbons. *Nat. Mater.* **8**, 235–242 (2009).
6. Ma, L., Wang, J. & Ding, F. Recent progress and challenges in graphene nanoribbon synthesis. *ChemPhysChem* **14**, 47–54 (2013).
7. Segawa, Y., Ito, H. & Itami, K. Structurally uniform and atomically precise carbon nanostructures. *Nat. Rev. Mater.* **1**, 15002 (2016).
8. Chen, L., Hernandez, Y., Feng, X. & Müllen, K. From nanographene and graphene nanoribbons to graphene sheets: chemical synthesis. *Angew. Chem. Int. Ed.* **51**, 7640–7654 (2012).
9. Narita, A., Feng, X. & Müllen, K. Bottom-up synthesis of chemically precise graphene nanoribbons. *Chem. Rec.* **15**, 295–309 (2015).
10. Yang, X. et al. Two-dimensional graphene nanoribbons. *J. Am. Chem. Soc.* **130**, 4216–4217 (2008).
11. Cai, J. et al. Atomically precise bottom-up fabrication of graphene nanoribbons. *Nature* **466**, 470–473 (2010).
12. Narita, A. et al. Synthesis of structurally well-defined and liquid-phase-processable graphene nanoribbons. *Nat. Chem.* **6**, 126–132 (2014).
13. Ruffieux, P. et al. On-surface synthesis of graphene nanoribbons with zigzag edge topology. *Nature* **531**, 489–492 (2016).
14. Talirz, L. et al. On-surface synthesis and characterization of 9-atom wide armchair graphene nanoribbons. *ACS Nano* **11**, 1380–1388 (2017).
15. Sakaguchi, H., Song, S., Kojima, T. & Nakae, T. Homochiral polymerization-driven selective growth of graphene nanoribbons. *Nat. Chem.* **9**, 57–63 (2017).
16. Jordan, R. S. et al. Synthesis of graphene nanoribbons via the topochemical polymerization and subsequent aromatization of a diacetylene precursor. *Chem* **1**, 78–90 (2016).
17. Jordan, R. S. et al. Synthesis of $N = 8$ armchair graphene nanoribbons from four distinct polydiacetylenes. *J. Am. Chem. Soc.* **139**, 15878–15890 (2017).
18. Yang, W., Lucotti, A., Tommasini, M. & Chalifoux, W. A. Bottom-up synthesis of soluble and narrow graphene nanoribbons using alkyne benzannulations. *J. Am. Chem. Soc.* **138**, 9137–9144 (2016).
19. Ito, H., Ozaki, K. & Itami, K. Annulative π-extension (APEX): an enabling reaction for rapid access to fused aromatics, heteroaromatics, and nanographenes. *Angew. Chem. Int. Ed.* **56**, 11144–11164 (2017).
20. Pouliot, J.-R., Grenier, F., Blaskovits, J. T., Beaupré, S. & Leclerc, M. Direct (hetero) arylation polymerization: simplicity for conjugated polymer synthesis. *Chem. Rev.* **116**, 14225–14274 (2016).
21. Teo, Y. C., Lai, H. W. H. & Xia, Y. Synthesis of ladder polymers: developments, challenges, and opportunities. *Chemistry* **23**, 14101–14112 (2017).
22. Ozaki, K., Kawasumi, K., Shibata, M., Ito, H. & Itami, K. One-shot K-region-selective annulative π-extension for nanographene synthesis and functionalization. *Nat. Commun.* **6**, 6251 (2015).
23. Yano, Y., Ito, H., Segawa, Y. & Itami, K. Helically twisted tetracene: synthesis, crystal structure, and photophysical properties of hexabenzo[*a,c,fg,j,l,op*] tetracene. *Synlett* **27**, 2081–2084 (2016).
24. Yokozawa, T. & Ohta, Y. Transformation of step-growth polymerization into living chain-growth polymerization. *Chem. Rev.* **116**, 1950–1968 (2016).
25. Huang, Y. et al. Poly(ethylene oxide) functionalized graphene nanoribbons with excellent solution processability. *J. Am. Chem. Soc.* **138**, 10136–10139 (2016).
26. Webster, O. W. Living polymerization methods. *Science* **251**, 887–893 (1991).
27. Ozaki, K. et al. One-step annulative π-extension of alkynes with dibenzosiloles or dibenzogermoles by palladium/*o*-chloranil catalysis. *Angew. Chem. Int. Ed.* **56**, 1361–1364 (2017).
28. Prezzi, D., Varsano, D., Ruini, A., Marini, A. & Molinari, E. Optical properties of graphene nanoribbons: the role of many-body effects. *Phys. Rev. B* **77**, 041404 (2008).
29. Vandescuren, M., Hermet, P., Meunier, V., Henrard, L. & Lambin, P. Theoretical study of the vibrational edge modes in graphene nanoribbons. *Phys. Rev. B* **78**, 195401 (2008).
30. Gillen, R., Mohr, M. & Maultzschm, J. Symmetry properties of vibrational modes in graphene nanoribbons. *Phys. Rev. B* **81**, 205426 (2010).

Acknowledgements This work was supported by the ERATO programme from JST (JPMJER1302 to K.I.), the JSPS KAKENHI (grants JP26810057, JP16H00907, JP17K1955 and JP18H02019 to H.I.), the SUMITOMO Foundation (141495 to H.I.) and the DAIKO Foundation (H.I.). We acknowledge Taoka Chemical Co., Ltd for providing samples. We thank K. Kuwata (Nagoya University) for MALDI–TOF mass measurements; M. Kamigaito and M. Uchiyama (Nagoya University) for measurements of molecular weight by size-exclusion chromatography multi-angle light scattering (SEC-MALS); T. Hashizume (Hitachi Ltd and Tokyo Institute of Technology) for advice on the STM and AFM measurements; C. M. Crudden (Queen's University and Nagoya University), G. J. P. Perry (Nagoya University) and A. Miyazaki (Nagoya University) for comments and proofreading. Computations were performed at the Research Center for Computational Science, Okazaki, Japan. ITbM is supported by the World Premier International Research Center (WPI) Initiative, Japan.

Reviewer information *Nature* thanks Lawrence T. Scott and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Author contributions K.I. and H.I. conceived the idea and directed the project. Y.Y., K. Matsushima and H.I. conducted the experiments and theoretical calculations. K. Matsui, A.T. and N.M. conducted the STM experiments. F.W. conducted the AFM experiments. Y.M. supervised the scanning probe microscopy experiments and provided advice on spectroscopic analysis. H.I. and K.I. prepared the manuscript with feedback from the other authors.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1331-z.

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1331-z.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to H.I. or K.I.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

General procedure for living APEX polymerization. The general experimental procedure of living APEX polymerization of **I** and **M** with a ratio of 1:100 (described in Fig. 1a) is as follows. Initiator **I** (0.40 mg, 2.25 μmol), monomer **M** (100 mg, 225 μmol), $Pd(OCOCF_3)_2$ (75.1 mg, 225 μmol), $AgSbF_6$ (77.6 mg, 450 μmol), *o*-chloranil (110 mg, 450 μmol) and 1,2-dichloroethane (2.3 ml) are added into a 5-ml Schlenk tube containing a magnetic stirring bar under a N_2 atmosphere. After stirring the mixture at 80 °C for 12 h, the reaction mixture is cooled to room temperature, passed through a short pad of silica gel and then washed with a metal scavenger with CH_2Cl_2. The solvent is removed in vacuo to give the crude product, and SEC analysis shows the formation of a polymer with $M_n = 3.2 \times 10^4$ Da and $Đ = 1.25$. Purification by SEC with chloroform as an eluent affords fjord-type GNR **2** as a black powder (88.0 mg, 85% yield). The living APEX polymerization reactions of **I** and **M** with ratios 1:10, 1:50, 1:300, 1:500 are conducted under the same reaction conditions to afford GNR **2** with $M_n = 2.9 \times 10^3$ Da ($Đ = 1.25$), 1.3×10^4 Da (1.23), 9.7×10^4 Da (1.22) and 1.5×10^5 Da (1.22), respectively.

General procedure for the transformation of fjord-type GNR 2 to armchair-type GNR 8. Fjord-type GNR **2** (30 mg, 30 μmol, $M_n = 3.2 \times 10^4$ Da and $Đ = 1.21$) and CH_2Cl_2 (120 ml) are added into a 200-ml two-necked round-bottom flask containing a magnetic stirring bar under a N_2 atmosphere. A suspension of iron(III) chloride (1.75 g, 10.8 mmol, 7 equiv. against hydrogen atoms on fjord-regions, estimated from M_n) in nitromethane (15 ml) is added to this solution. After stirring the mixture at room temperature for 72 h, the reaction is quenched by the addition of methanol to form a dark red precipitate. Filtration by suction and intensive washing with methanol give armchair-type GNR **8** as a dark red powder (28.5 mg, 98% yield). The SEC analysis shows the formation of GNR **8** with $M_n = 2.4 \times 10^4$ Da and $Đ = 1.32$.

Determination of M_n, M_w and $Đ$ of synthesized GNRs. The molecular weight (M_n and M_w) and $Đ$ (M_w/M_n) of all synthesized GNRs were measured by SEC using a Shimadzu Prominence 2000 instrument equipped with two in-line linear polystyrene gel columns (TOSOH TSKgel Multipore H_{XL}-M SEC columns 7.8 mm × 300 mm) at 40 °C, and tetrahydrofuran containing 0.1 wt% tetra-*n*-butylammonium bromide was used as the eluent at a flow rate of 1.0 ml min^{-1}. The molecular weight calibration curve was obtained with standard PS (TOSOH TSKgel PS standard).

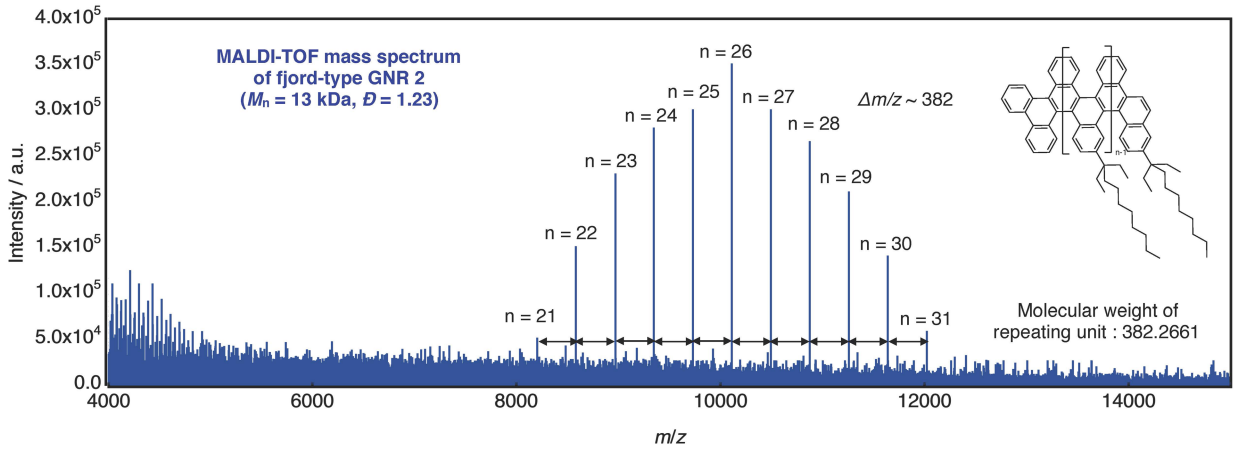
STM characterization of GNRs 2 and 8. A low-temperature STM (Unisoku, USS-10092LSC) was used to characterize the morphology of the GNR samples prepared on HOPG. The sample substrates were prepared by drop-casting solutions of GNR **2** (~0.01 mg mL^{-1}, $M_n = 1.5 \times 10^5$ Da, $Đ = 1.22$) or of **8** (~0.0015 mg mL^{-1}, $M_n = 2.4 \times 10^4$ Da, $Đ = 1.32$) in 1,2,4-trichlorobenzene on HOPG; the solutions were heated at 363 K, followed by evaporation of 1,2,4-trichlorobenzene at 363 K under air. STM images were taken in constant current mode under ultrahigh vacuum at 78 K (liquid N_2 cooling).

AFM characterization of GNR 2. A solution of about 0.01 mg mL^{-1} fjord-type GNR **2** with $M_n = 3.4 \times 10^4$ Da and $Đ = 1.21$ in 1,2,4-trichlorobenzene was drop-casted on the HOPG substrate, which was heated on a hot plate at 393 K under air. After settling for several tens of seconds, the solvent was removed by blowing N_2 gas in one direction and further dried for 3 min at 393 K. AFM (Bruker, DimensionFastScan) images were taken in tapping mode in air at room temperature.

Data availability

The datasets generated and/or analysed during the current study are available from the corresponding authors on reasonable request.

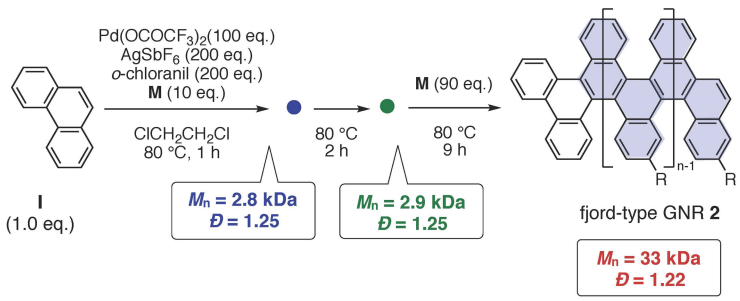
a



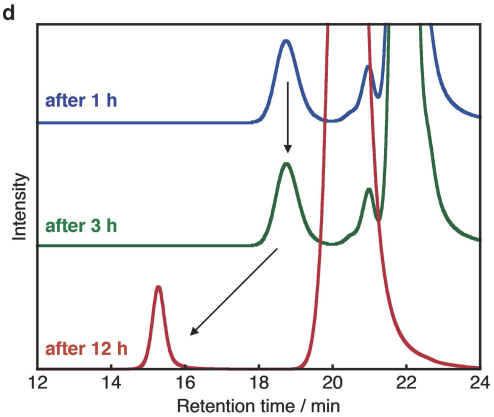
b

m/z	n										
	21	22	23	24	25	26	27	28	29	30	31
calcd for M^+	8205.6663	8587.9324	8970.1985	9352.4646	9734.7307	10116.9968	10449.2629	10881.5390	11263.7951	11636.0612	12028.3273
found for M^+	8205.6667	8587.9359	8970.2478	9352.4592	9734.6991	10116.9099	10449.2512	10881.5167	11263.7934	11636.0116	12028.2843

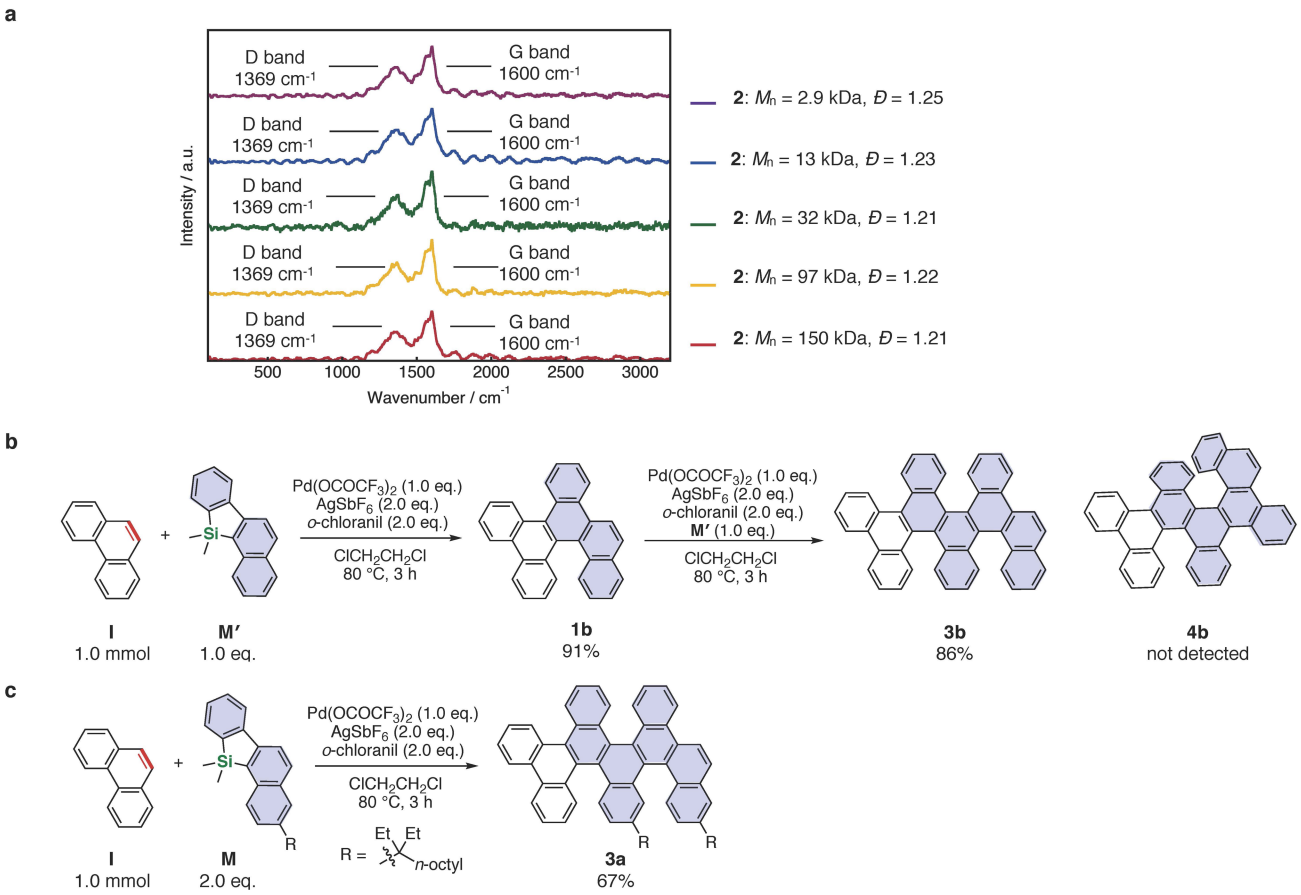
c



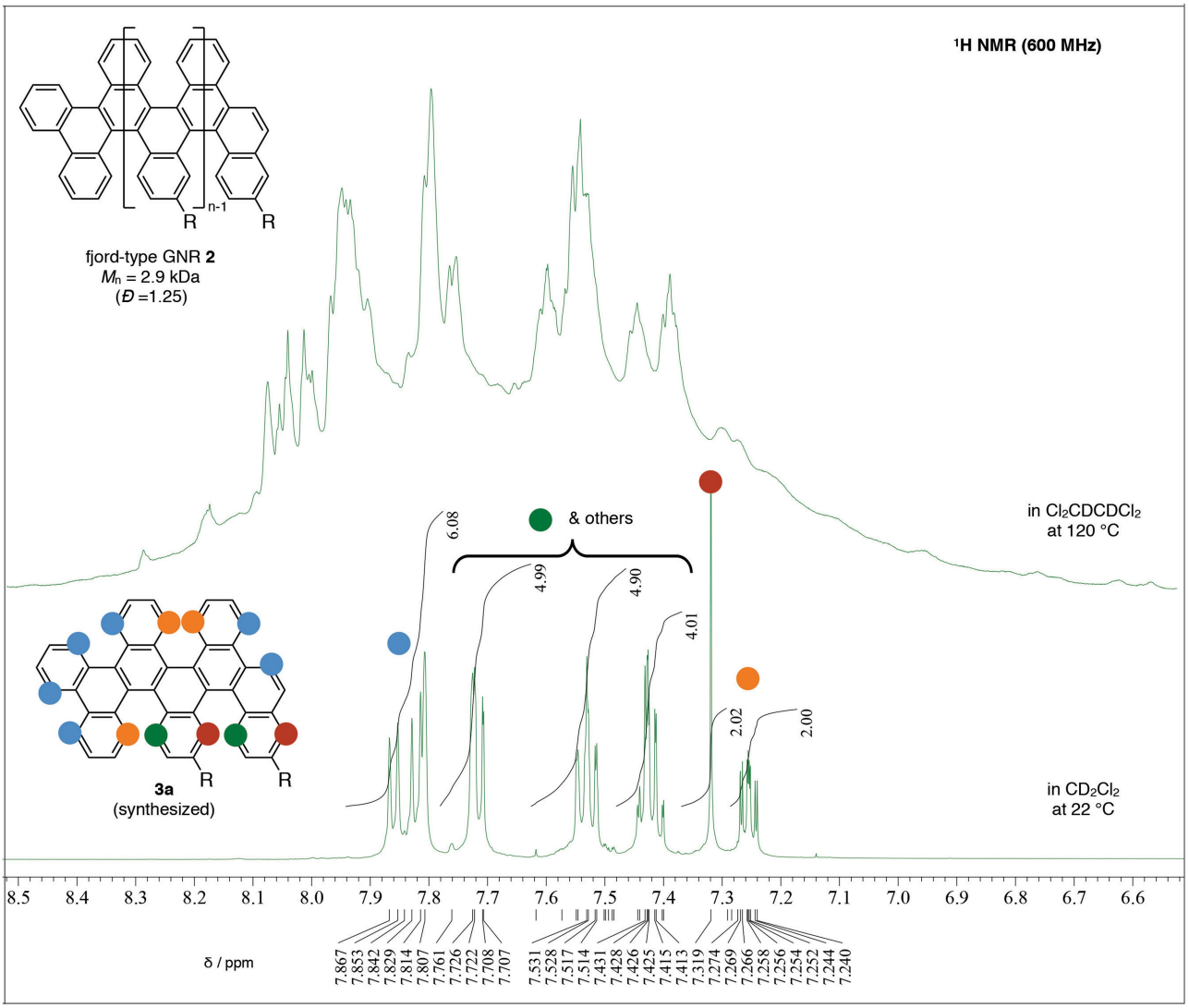
d



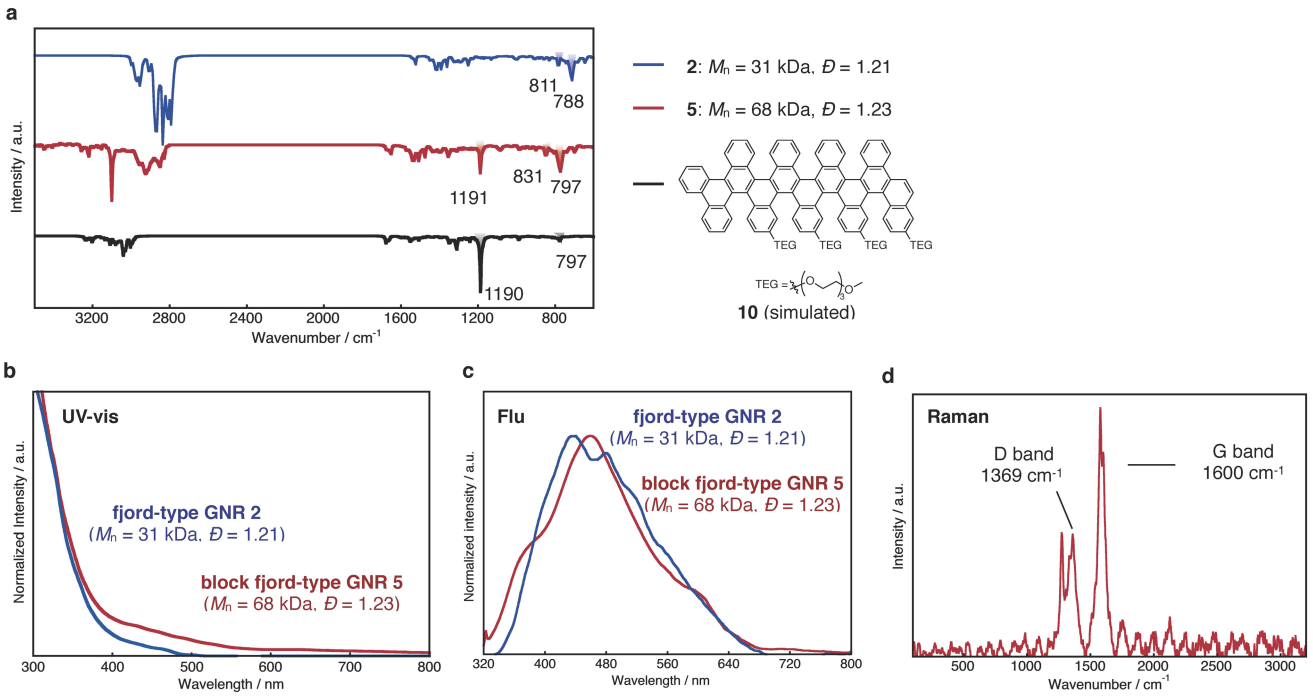
Extended Data Fig. 1 | Mass spectroscopy analysis and confirmation of livingness of APEX polymerization. a, MALDI–TOF mass spectrum (reflection mode) of fjord-type GNR **2** ($M_n = 1.3 \times 10^4$, $Đ = 1.23$) with *trans*-2-[3-(4-*tert*-butylphenyl)-2-methyl-2-propenylidene]malononitrile (DCTB) as the matrix. **b**, Summary of observed and calculated molecular ion peaks. **c**, Confirmation of livingness of APEX polymerization by one-pot SEC analyses. **d**, SEC charts of the reaction progress after 1 h, 3 h and 12 h.



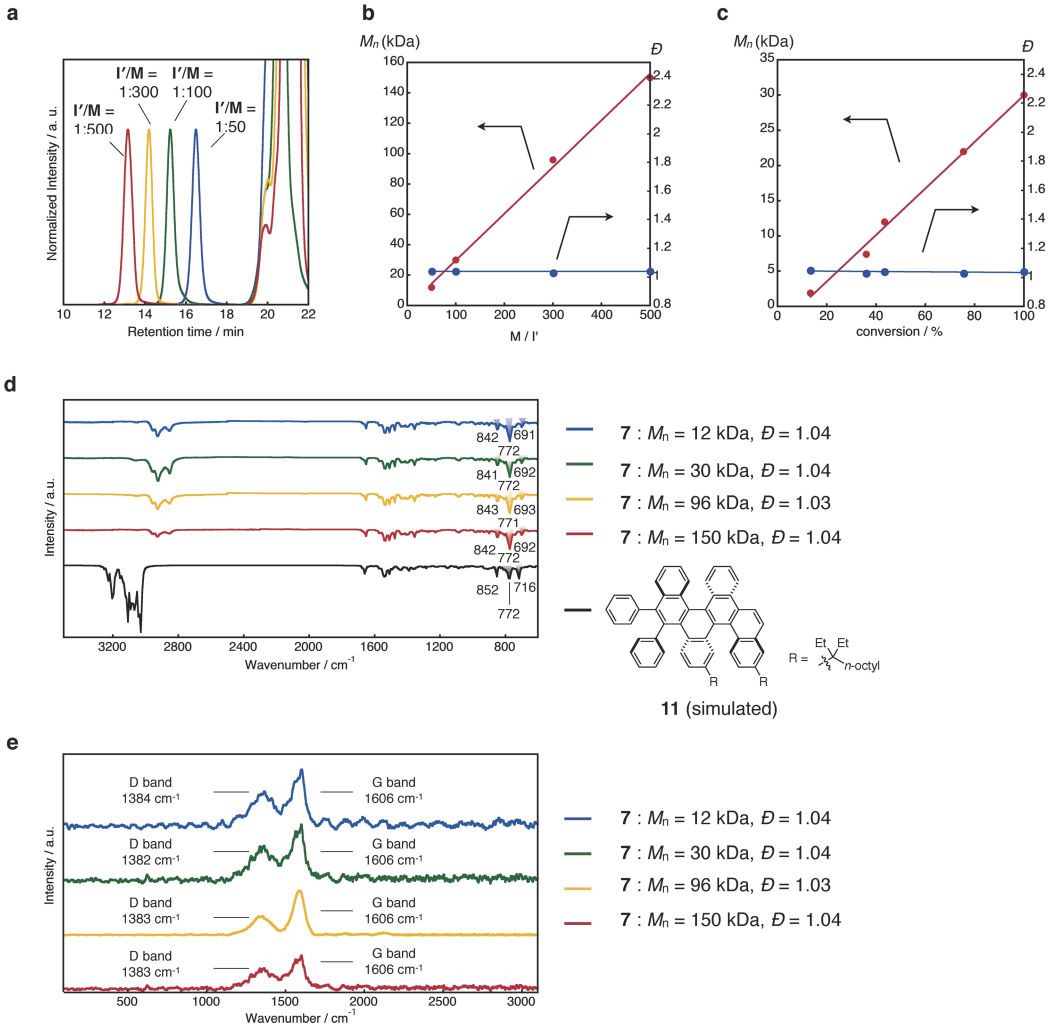
Extended Data Fig. 2 | Raman spectroscopy analyses of fjord-type GNR 2 and synthesis of model dimer. a, Raman spectra of fjord-type GNR **2**, obtained with a 532-nm excitation laser. **b**, Synthesis of model dimer **3b** by step-by-step APEX reaction of phenanthrene. **c**, Synthesis of model dimer **3a** by APEX reaction.



Extended Data Fig. 3 | ^{1}H NMR spectroscopy analysis of fjord-type GNR 2. Observed ^{1}H NMR spectra of GNR **2** (in $Cl_2CDCDCl_2$ at 120 °C) and dimer **3a** (in CD_2Cl_2 at 22 °C). All of the peaks of **3a** range from 7.2 to 7.9 p.p.m. Two doublet-of-doublet peaks appear at a higher magnetic field (chemical shift δ = 7.2–7.3 p.p.m.), which are attributed to the most shielded aromatic hydrogen atoms, in the orange-coloured fjord regions. Other aromatic hydrogen atoms in the fjord regions next to the alkyl substituents (red circle) are found as a singlet peak at 7.32 p.p.m. On the other hand, the most deshielded aromatic hydrogen atoms in the bay and cove regions appear at 7.8–7.9 p.p.m. The ^{1}H NMR spectrum of fjord-type GNR **2** shows broadened peaks at 7.0–8.3 p.p.m., which is expected to reflect the repeating [5]helicene structure.

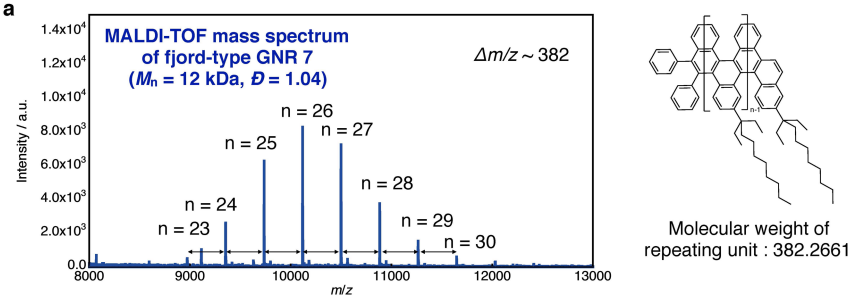


Extended Data Fig. 4 | Comparison of IR, absorption, emission and Raman spectroscopy of 2 and 5. a, Observed IR spectra of GNRs **2** and **5** (blue and red lines) and IR spectrum of tetramer **10** (black line) predicted by DFT calculations at the B3LYP/6-31G(d) level of theory. **b**, UV-vis absorption spectra of GNRs **2** and **5** (blue and red lines). **c**, Emission spectra of GNRs **2** and **5** (blue and red lines) excited at 320 nm. **d**, Raman spectra of GNR **5**, obtained by excitation by a 532-nm laser.



Extended Data Fig. 5 | Analytical SEC, IR and Raman spectroscopy data for fjord-type GNR 7. a, SEC charts of the crude product of reactions with various **I′/M** ratios. **b**, **c**, M_n and M_w profiles versus **M/I′** and conversion of the monomer in the living APEX polymerization. **d**, Observed IR spectra of GNR **7** with different polymer lengths (blue, green, yellow and red lines) and IR spectrum of dimer **11** (black line) predicted by DFT calculations at the B3LYP/6-31G(d) level of theory. **e**, Raman spectra of GNR **7** with different polymer lengths (blue, green, yellow and red lines), obtained with excitation by a 532 nm laser.

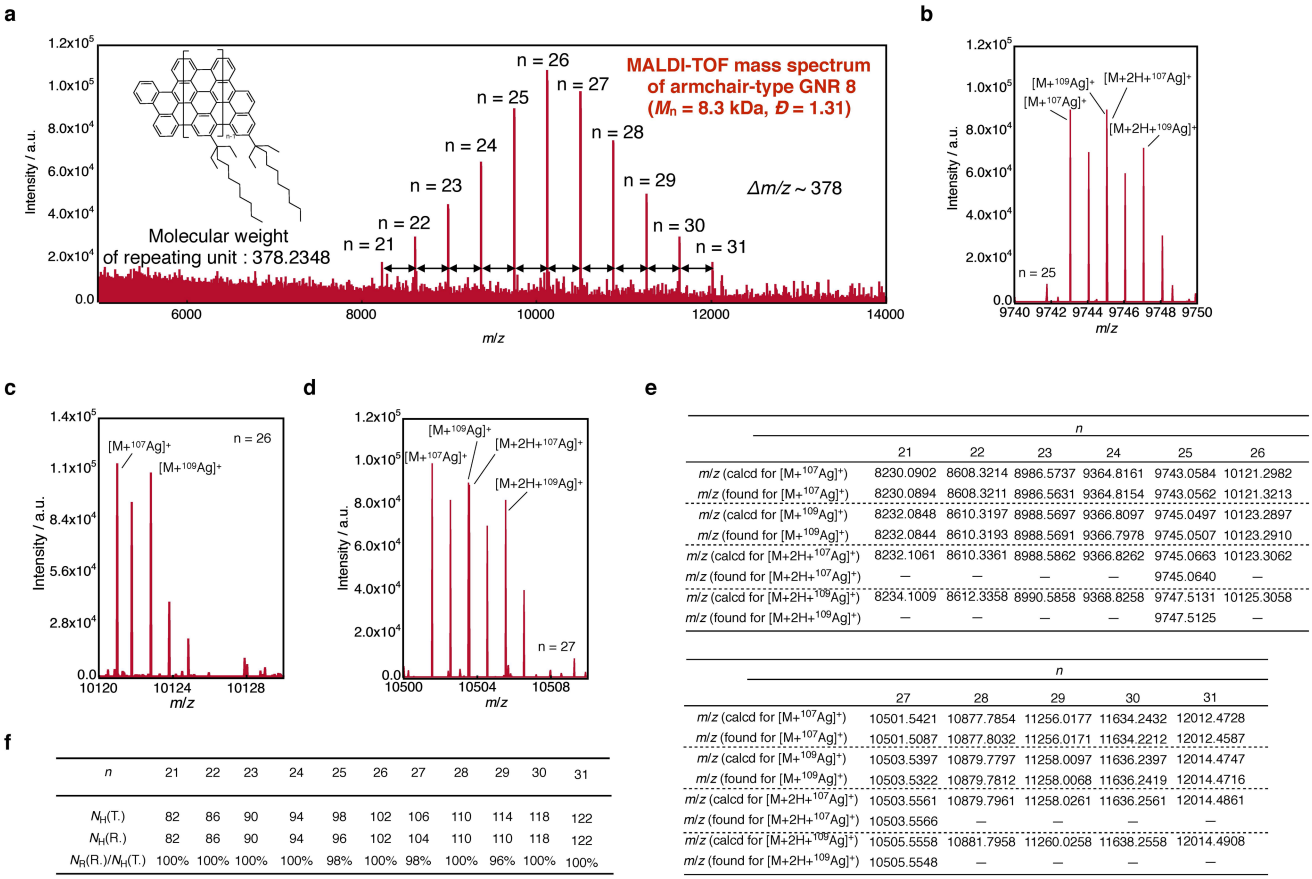
a



b

	n							
	23	24	25	26	27	28	29	30
m/z (calcd for M^+)	8972.2142	9354.4803	9736.7465	10119.0125	10501.2786	10883.5447	11265.8108	11648.0769
m/z (found for M^+)	8972.7060	9354.3802	9736.7455	10119.7223	10501.4954	10883.3387	11265.2346	11648.2012

Extended Data Fig. 6 | Analytical MALDI–TOF data for fjord-type GNR 7. a, MALDI–TOF mass spectrum (reflection mode) of fjord-type GNR **7** ($M_n = 1.2 \times 10^4$ Da, $Đ = 1.04$) with DCTB as the matrix. **b**, Summary of observed and calculated molecular ion peaks.



e

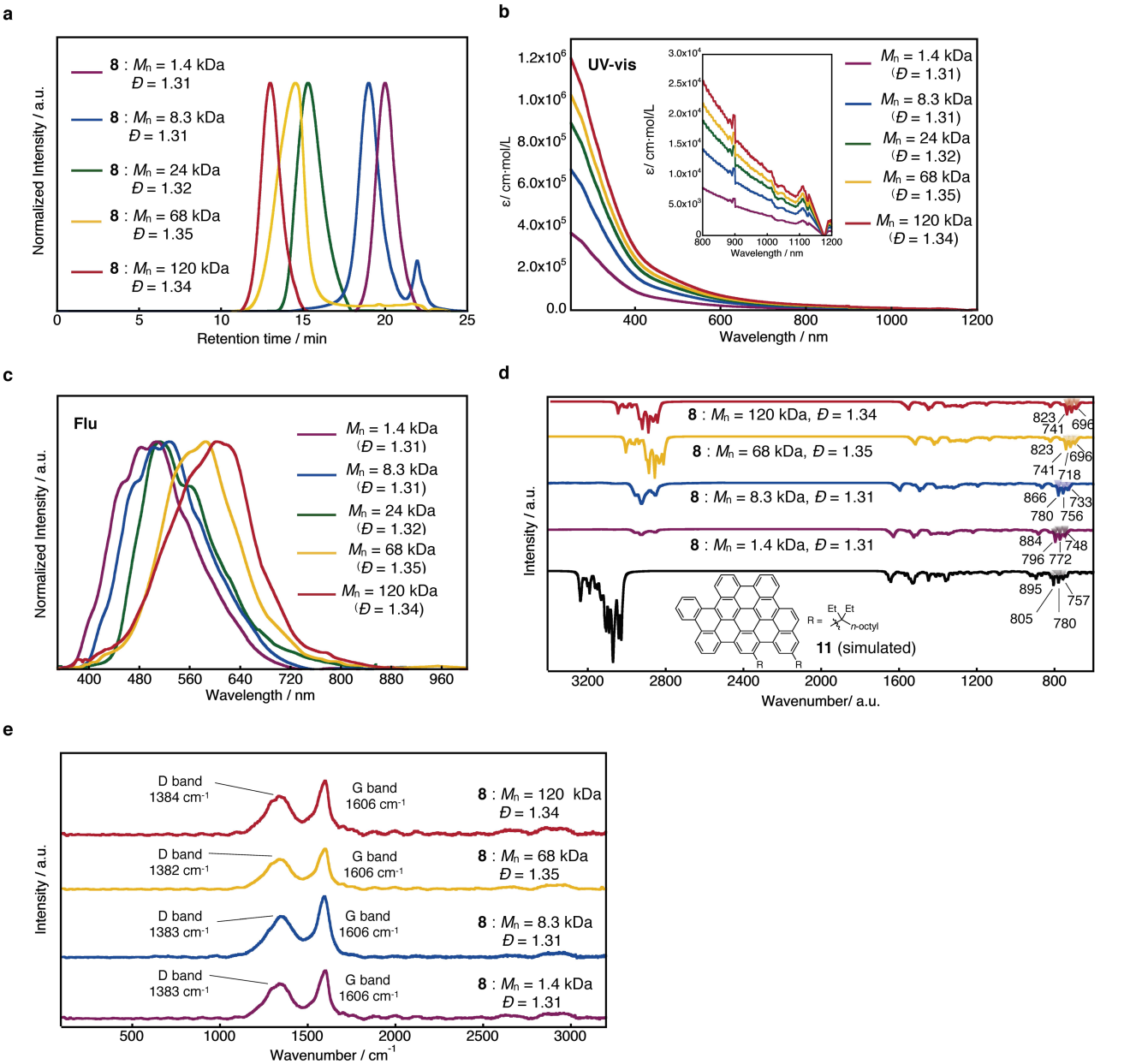
	n					
	21	22	23	24	25	26
m/z (calcd for $[M+^{107}Ag]^+$)	8230.0902	8608.3214	8986.5737	9364.8161	9743.0584	10121.2982
m/z (found for $[M+^{107}Ag]^+$)	8230.0894	8608.3211	8986.5631	9364.8154	9743.0562	10121.3213
m/z (calcd for $[M+^{109}Ag]^+$)	8232.0848	8610.3197	8988.5697	9366.8097	9745.0497	10123.2897
m/z (found for $[M+^{109}Ag]^+$)	8232.0844	8610.3193	8988.5691	9366.7978	9745.0507	10123.2910
m/z (calcd for $[M+2H+^{107}Ag]^+$)	8232.1061	8610.3361	8988.5862	9366.8262	9745.0663	10123.3062
m/z (found for $[M+2H+^{107}Ag]^+$)	—	—	—	—	9745.0640	—
m/z (calcd for $[M+2H+^{109}Ag]^+$)	8234.1009	8612.3358	8990.5858	9368.8258	9747.5131	10125.3058
m/z (found for $[M+2H+^{109}Ag]^+$)	—	—	—	—	9747.5125	—

	n				
	27	28	29	30	31
m/z (calcd for $[M+^{107}Ag]^+$)	10501.5421	10877.7854	11256.0177	11634.2432	12012.4728
m/z (found for $[M+^{107}Ag]^+$)	10501.5087	10877.8032	11256.0171	11634.2212	12012.4587
m/z (calcd for $[M+^{109}Ag]^+$)	10503.5397	10879.7797	11258.0097	11636.2397	12014.4747
m/z (found for $[M+^{109}Ag]^+$)	10503.5322	10879.7812	11258.0068	11636.2419	12014.4716
m/z (calcd for $[M+2H+^{107}Ag]^+$)	10503.5561	10879.7961	11258.0261	11636.2561	12014.4861
m/z (found for $[M+2H+^{107}Ag]^+$)	10503.5566	—	—	—	—
m/z (calcd for $[M+2H+^{109}Ag]^+$)	10505.5558	10881.7958	11260.0258	11638.2558	12014.4908
m/z (found for $[M+2H+^{109}Ag]^+$)	10505.5548	—	—	—	—

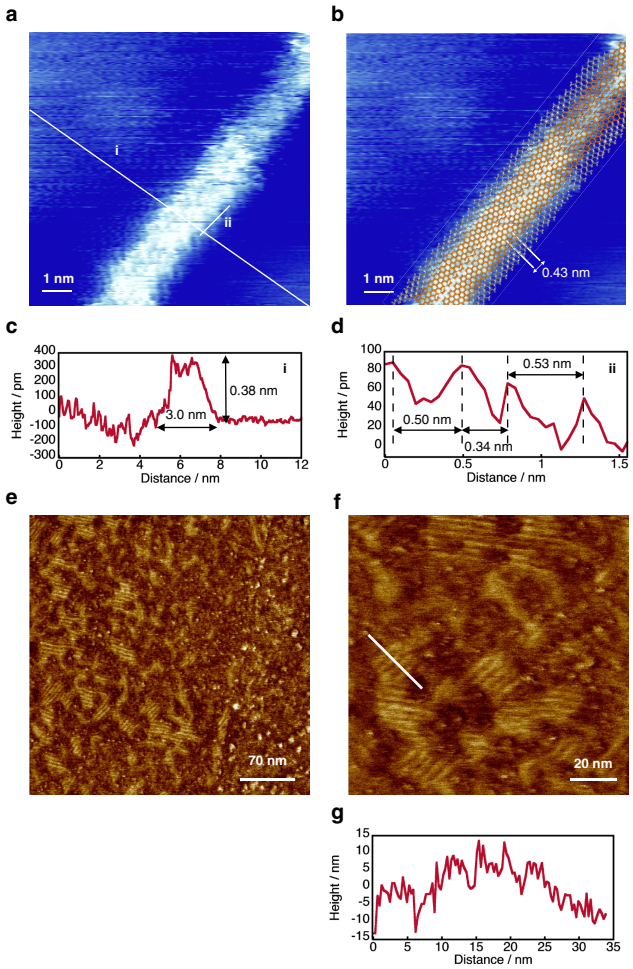
f

n	21	22	23	24	25	26	27	28	29	30	31
N_H(T.)	82	86	90	94	98	102	106	110	114	118	122
N_H(R.)	82	86	90	94	96	102	104	110	110	118	122
N_R(R.)/N_H(T.)	100%	100%	100%	100%	98%	100%	98%	100%	96%	100%	100%

Extended Data Fig. 7 | Formation of armchair-type GNR 8. **a**–**d**, MALDI–TOF mass spectra (reflection mode) of armchair-type GNR **8** ($M_n = 8.3 \times 10^3$ Da, $Đ = 1.31$) with 7,7,8,8-tetracyanoquinodimethane as the matrix and $AgOCOCF_3$ as an additive. **e**, Summary of observed and calculated molecular ion peaks. The formation of GNR **8** with $M_n = 2.4 \times 10^4$ Da ($Đ = 1.32$) was confirmed by MALDI–TOF mass spectroscopy analysis, where 11 distinct mass peaks, each with gaps of $m/z = 378$ (DP(n) = 21–31), were observed. Around the $[M+Ag]^+$ peaks with $n = 25$ and 27, mass peaks derived from two additional hydrogen atoms ($[M+Ag+2H]^+$) were clearly observed, which indicates partial completion of the Scholl reaction. **f**, Efficiency of the transformation reaction to the armchair-type GNR **8**[10]. According to the literature[10], we also estimated the efficiency of cyclodehydrogenation at each observed mass peak by using the theoretical number of lost hydrogen atoms during cyclodehydrogenation (N_H(T.)) and the number of lost hydrogen atoms derived from MALDI–TOF mass measurements (N_H(R.)) As a result, 96%–100% efficiencies (N_H(T.)/N_H(R.)) were confirmed at each m/z peak. The highest peak was found at $m/z = 10{,}121.3213$ (DP(n) = 26, $[M+^{107}Ag]^+$), which is higher than the PS-based M_n, but the top peak DP (DP_{top}) and the dispersity pattern were almost identical to those of the starting fjord-type GNR **2**. No peak corresponding to shorter oligomers or fragments was observed in the full-scale MALDI–TOF mass spectrum (Supplementary Fig. 19). The detection of larger GNRs with a molecular weight of over 20,000 Da appeared to be difficult owing to limitations in the measurement of the highly aggregating flat armchair GNR in the MALDI–TOF mass instrument.



Extended Data Fig. 8 | Analytical SEC, UV-vis absorption, emission, IR and Raman spectroscopy data for armchair-type GNR 8. a, SEC charts of armchair-type GNR **8** obtained by the Scholl reaction of fjord-type GNR **2** with different polymer lengths. **b**, UV-vis absorption spectra of **8. c**, Emission spectra of **8**. **d**, Observed IR spectra of GNR **8** with different polymer lengths (red, yellow, blue and purple lines) and IR spectrum of dimer **11** (black line) predicted by DFT calculations at the B3LYP/6-31G(d) level of theory. **e**, Raman spectra of GNR **8** with different polymer lengths (red, yellow, blue and purple lines), obtained by excitation by a 532-nm laser.



Extended Data Fig. 9 | STM and AFM images of armchair-type GNR 8. **a**, STM image of armchair-type GNR **8** ($M_n = 2.4 \times 10^4$ Da, $Đ = 1.32$) deposited on HOPG ($U = 1.0$ V, $I = 250$ pA, $T = 78$ K). **b**, Dimeric assembly model of GNR **8**. **c**, **d**, Cross-sectional height profile taken perpendicularly (i) and horizontally (ii) with respect to the wire shown in **a**. The observed substance had about 0.4 nm thickness and 3 nm width, and the repeating peripheral stripes had 0.4–0.5 nm intervals ((i) and (ii)). The existence of this substance was considered as the assembly of two molecules of **8**, as depicted in **b**, and the observed periodicity of 0.4–0.5 nm corresponds to the longitudinal length between the alkyl side chains ((ii) in **d**). **e**, AFM phase image of self-assembling GNR **8** ($M_n = 2.4 \times 10^4$ Da, $Đ = 1.32$) on graphene cleaved on SiO_2. **f**, Magnified AFM phase image of **e**. **g**, Cross-sectional height profile taken along the white line in **f**, perpendicularly to the observed stripe.

LETTER

https://doi.org/10.1038/s41586-019-1349-2

Correcting datasets leads to more homogeneous early-twentieth-century sea surface warming

Duo Chan[1], Elizabeth C. Kent[2], David I. Berry[2] & Peter Huybers[1]*

Existing estimates of sea surface temperatures (SSTs) indicate that, during the early twentieth century, the North Atlantic and northeast Pacific oceans warmed by twice the global average, whereas the northwest Pacific Ocean cooled by an amount equal to the global average[1–4]. Such a heterogeneous pattern suggests first-order contributions from regional variations in forcing or in ocean–atmosphere heat fluxes[5,6]. These older SST estimates are, however, derived from measurements of water temperatures in ship-board buckets, and must be corrected for substantial biases[7–9]. Here we show that correcting for offsets among groups of bucket measurements leads to SST variations that correlate better with nearby land temperatures and are more homogeneous in their pattern of warming. Offsets are identified by systematically comparing nearby SST observations among different groups[10]. Correcting for offsets in German measurements decreases warming rates in the North Atlantic, whereas correcting for Japanese measurement offsets leads to increased and more uniform warming in the North Pacific. Japanese measurement offsets in the 1930s primarily result from records having been truncated to whole degrees Celsius when the records were digitized in the 1960s. These findings underscore the fact that historical SST records reflect both physical and social dimensions in data collection, and suggest that further opportunities exist for improving the accuracy of historical SST records[9,11].

According to recent estimates from the National Oceanic and Atmospheric Administration (NOAA)[1], global average SST warmed by 0.43 °C between 1908 and 1941. Whereas the North Atlantic warmed by 0.82 °C, the North Pacific showed a bimodal structure, with the northwest Pacific cooling by −0.39 °C and the northeast Pacific warming by 1.02 °C. Other gridded SST products give similarly disparate SST trends for the early twentieth century (Table 1 and Extended Data Fig. 1), and together these SST estimates suggest that internal modes of variability strongly contributed to early-twentieth-century climate change. Specifically, the Atlantic Multidecadal Oscillation and the Pacific Decadal Oscillation have been suggested to account for regional variations as well as more than half of the global warming between 1908 and 1941 (refs [6,12]). Model simulations of the atmospheric and oceanic response to prescribed radiative forcing do not, however, reproduce either the magnitude[13,14] or the pattern[5,15] of the early-twentieth-century warming seen in observations (Extended Data Fig. 1e). Difficulty in reproducing observations has been suggested to arise from deficiencies in how radiative forcing is prescribed[16] or from model limitations in representing internal climate variability[17,18].

Another possibility is that observational estimates of SST changes contain undetected biases, for which there are some precedents. Difficulty in simulating a slowdown in global warming between 1997 and 2012 was partly reconciled by revising SST estimates[19], amongst other considerations[20]. In another study[21], a jump in global temperature by 0.3 °C in 1945 was attributed to offsets between engine-room intake and bucket SST estimates.

The four major SST products covering the early twentieth century each rely upon the International Comprehensive Ocean-Atmosphere Data Set (ICOADS)[22], whose latest release is 3.0. It is estimated that 94% of observations between 1908 and 1941 were from buckets (Fig. 1). Bucket measurements of SST are biased by evaporative, sensible and solar heat fluxes that depend on a range of factors, including weather, ship deck height and bucket type[7]. For example, a canvas bucket left on deck for three minutes under typical wind and other weather conditions can give water temperatures that are approximately 0.5 °C cooler than a wooden bucket measured using the same protocol[7,9].

Previous corrections for bucket-measurement biases have involved assumptions that these biases change smoothly in space or time[1,4].

Table 1 | Early-twentieth-century SST trends

	ICOADSa	ICOADSb	ERSST5	COBESST2	HadISST2	HadSST3
Global trend	0.51 ± 0.03	0.56 ± 0.10	0.43	0.48	0.44 ± 0.04	0.47 ± 0.03
N Atlantic trend	0.85 ± 0.03	0.66 ± 0.11	0.82	0.79	0.74 ± 0.04	0.71 ± 0.03
N Pacific trend	0.31 ± 0.03	0.56 ± 0.11	0.37	0.39	0.32 ± 0.04	0.37 ± 0.03
NW Pacific trend	−0.35 ± 0.04	−0.02 ± 0.11	−0.39	−0.14	−0.34 ± 0.06	−0.30 ± 0.04
NE Pacific trend	0.86 ± 0.04	1.03 ± 0.12	1.02	0.85	0.85 ± 0.06	0.94 ± 0.04
East Asia trend	−0.06 ± 0.04	0.34 ± 0.11	−0.31	0.04	−0.09 ± 0.04	0.00 ± 0.04
Eastern US trend	0.92 ± 0.05	0.65 ± 0.13	0.75	0.82	0.61 ± 0.05	0.56 ± 0.05
PDO trend	1.82 ± 0.19	1.53 ± 0.20	2.03 ± 0.15	1.64 ± 0.11	2.02 ± 0.12	1.98 ± 0.18
East Asia air-temperature correlation	0.67	0.85	0.53	0.68	0.69	0.71
Eastern US air-temperature correlation	0.65	0.70	0.67	0.56	0.72	0.76
Correction pattern correlation	−0.49	−0.10	−0.49	−0.43	−0.45	−0.42

Trends are averaged over non-grey areas as in Fig. 3a, with the North Atlantic and North Pacific defined as poleward of 20° N; the northwest Pacific between 120° E and 180° E, and 25° N and 45° N; and the northeast Pacific between 120° W and 160° W, and 20° N and 60° N. East Asia and Eastern US regions are shown in Extended Data Fig. 5c. All trends are between 1908 and 1941, and are in units of °C per 34 years, with uncertainties reported at the 2 s.d. level. Each reported SST trend uncertainty, derived from dataset ensembles where available, includes contributions from bucket corrections, but only ICOADSb also accounts for groupwise offsets. Cross-correlations (Pearson's *r*) are reported between interannual air temperatures[26] and SSTs from East Asia and the Eastern US between 1908 and 1941, and between the spatial patterns of trends in SSTs (Fig. 3a and Extended Data Fig. 1a–d) and groupwise corrections (Fig. 3b).

[1]Department of Earth and Planetary Sciences, Harvard University, Cambridge, MA, USA. [2]National Oceanography Centre, Southampton, UK. *e-mail: phuybers@fas.harvard.edu

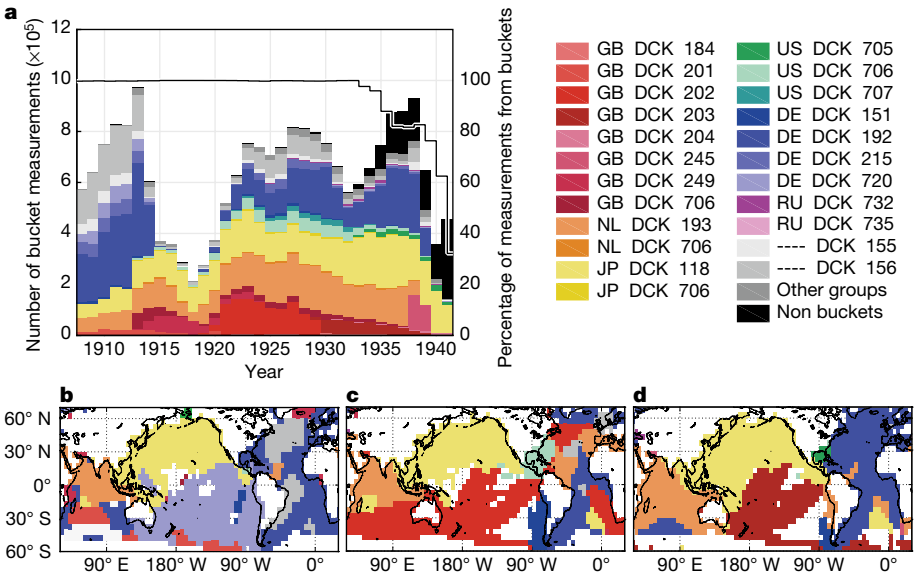


Fig. 1 | Groups of bucket SST measurements. a, Left-hand *y*-axis: number of bucket SST measurements from individual groups identified by country and deck information in ICOADS3.0. Country name abbreviations are: DE, Germany; GB, Great Britain; JP, Japan; NL, The Netherlands; RU, Russia; US, United States; and —, missing. Groups having fewer than 100,000 measurements are labelled as 'other groups'. Decks 118 and 762 are combined into 'JP DCK 118' because they are both Japanese Kobe Collection decks. Right-hand *y*-axis: percentage of measurements that have come from buckets, showing that nearly all observations before 1935 are from buckets (black line). **b**–**d**, Maps indicating nations that contribute the most observations within 5° × 5° grids for the periods 1908–1918 (**b**), 1919–1928 (**c**) and 1929–1941 (**d**). White grid boxes have fewer than three years of data.

HadSST3, for example, represents transitions between wooden buckets and less-insulated canvas buckets by using globally uniform, linear weights[4]. Another method assumes that differences between SST and night-time marine air temperatures remain similar to an average for 1971–2000 (ref. [1]). Simple assumptions regarding bucket corrections are commensurate with the limited metadata available for determining observational characteristics[2,4,7], but are questionable on the basis that ICOADS contains observations from a wide variety of ships that sailed for different purposes and sampled the ocean unevenly[22] (Fig. 1b–d).

More comprehensive bucket corrections are made possible by calculating offsets between groups of SST measurements that are nearby in space and time[10]. Specifically, we difference bucket SST measurements

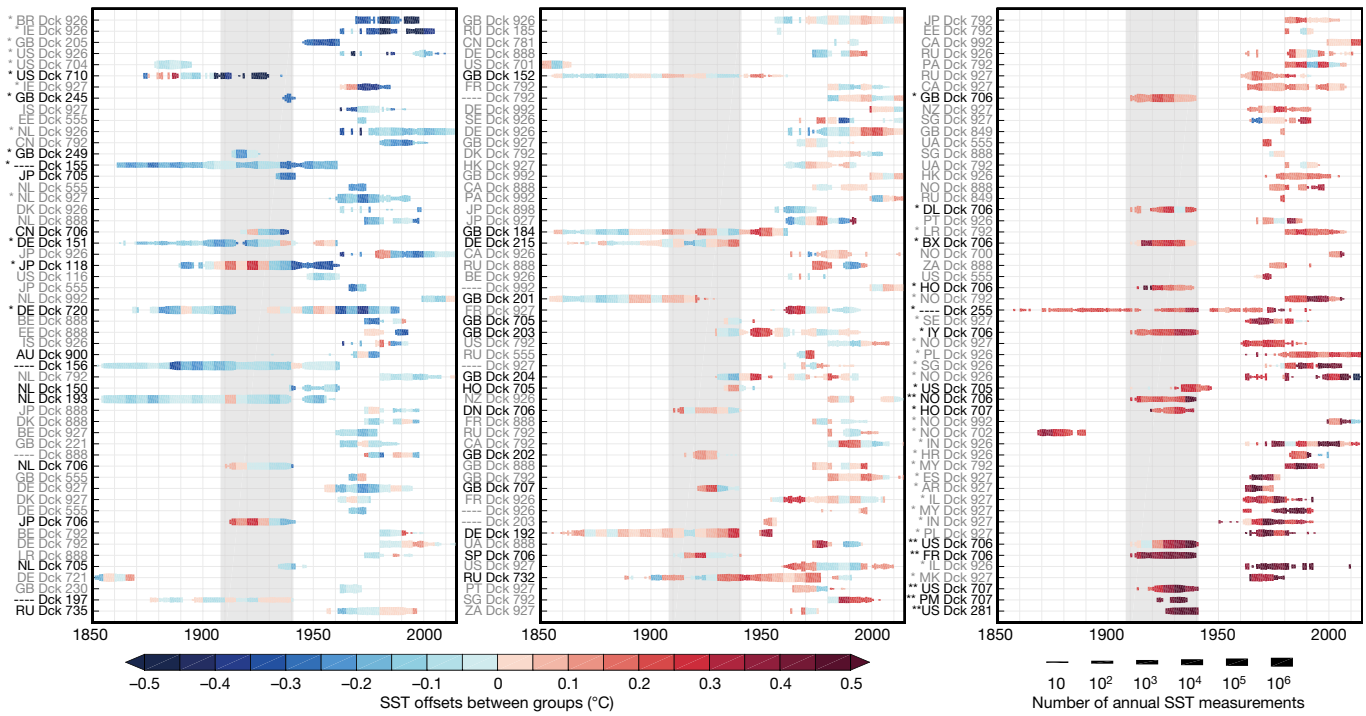


Fig. 2 | Relative offsets between groups of bucket SSTs. Groups for which fixed effects differ significantly from zero are indicated by * ($P < 0.05$). The 46 out of 162 groups that contribute data between 1908 and 1941 are indicated in black, and those remaining significant after a Bonferroni correction ($P < 0.05/46$) are indicated by **. Shading indicates the sum of fixed and five-yearly effects (regional effects are not shown). Bar widths indicate the number of SST measurements contributed by each group for each year. Abbreviated country names correspond to those in Fig. 1 and are listed in Supplementary Table 1.

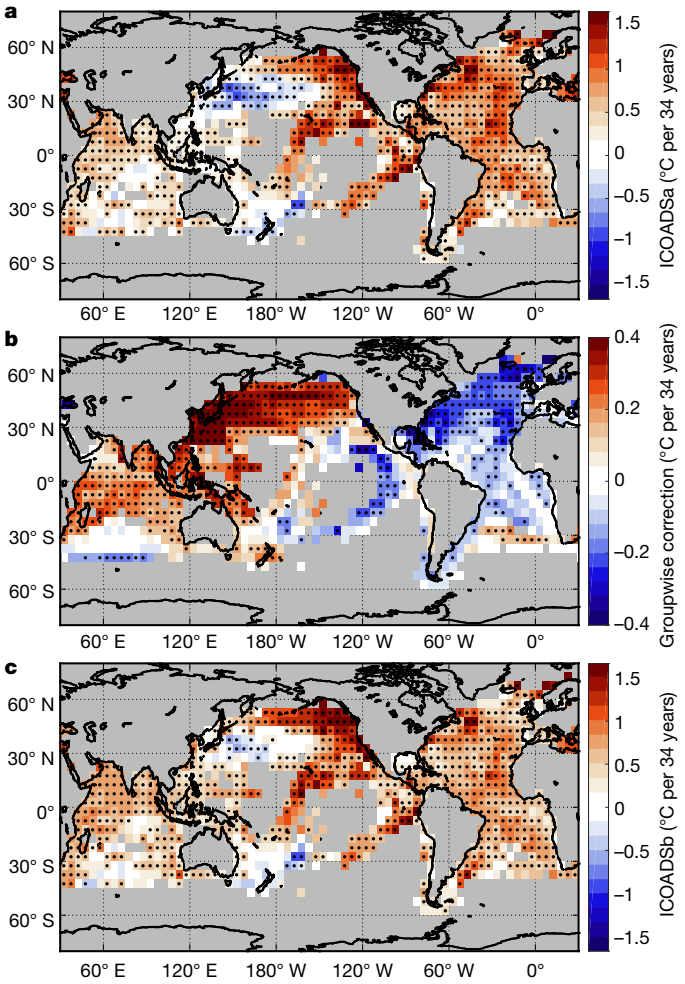


Fig. 3 | Maps of SST trends and corrections between 1908 and 1941. **a**, SST trends in ICOADSa are similar to patterns found in existing SST estimates (Extended Data Fig. 1). **b**, Trends associated with the corrections for groupwise offsets. Note that panel **b** is plotted on a different colour scale. **c**, SST trends in ICOADSb after applying groupwise corrections. Areas in grey are inadequately sampled for purposes of calculating trends (see Methods). Dots indicate significant trends ($P < 0.05$). In ICOADSb (**c**), 77% of boxes show statistically significant warming, whereas only 2% show significant cooling. By contrast, in ICOADSa (**a**), 6% of boxes indicate significant cooling.

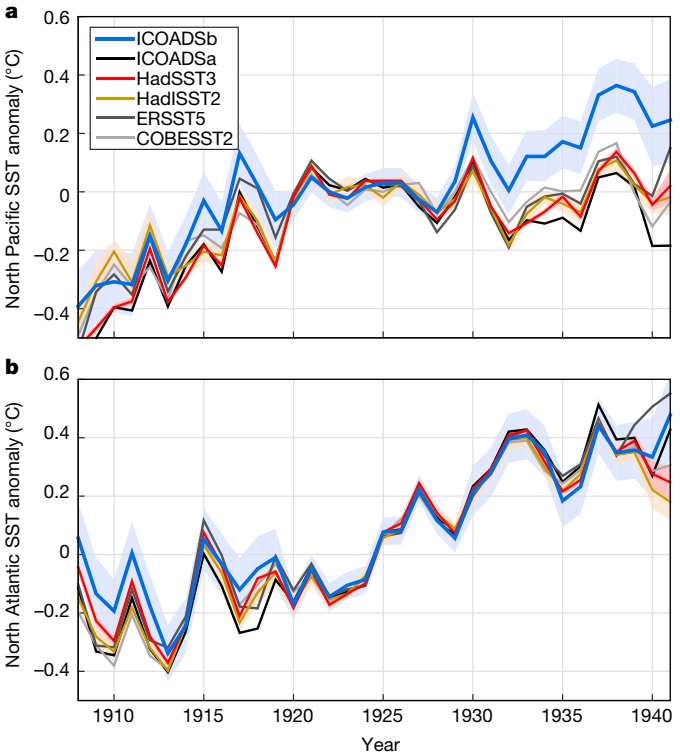


Fig. 4 | Diverging estimates of regional temperature variations. **a**, **b**, Annual SST anomalies from different datasets in the North Pacific (**a**) and North Atlantic (**b**) oceans. Anomalies are relative to the 1920–1929 mean of each SST estimate. ICOADSb shows greater warming in the North Pacific and less warming in the North Atlantic relative to previous estimates. Uncertainties associated with ICOADSb (blue shading, 2 s.d.) are for annual average SSTs for each sub-basin, and are an order of magnitude larger than those reported for HadSST3 (red shading). Note that those uncertainties included in HadSST3 are mostly removed when computing the anomaly. The discrepancy in annual average SST uncertainties is larger than the discrepancy for trends (Table 1).

that come from distinct groups of ships within 300 km and 2 days of one another, giving a dataset of 6.1×10^6 SST comparisons between 1908 and 1941. Groups are designated according to nations and 'decks', with the latter term inherited from the fact that marine observations were stored using decks of punch cards. SST differences are analysed using a linear-mixed-effects (LME) methodology after accounting for climatological effects associated with location, day of year, and hour of day[10]. Mean offsets between groups of SST measurements range from −0.3 °C to +0.6 °C (Supplementary Table 1). Of the 46 nation–deck groups that contribute SST observations between 1908 and 1941, 21 have significant offsets ($P < 0.05$), and 6 remain significant after applying a Bonferroni correction[23] for multiple hypothesis testing ($P < 0.05/n$, $n = 46$; Fig. 2).

The presence of systematic offsets between groups of measurements, combined with changes in the distribution of these groups over time (Fig. 1), is liable to introduce spurious SST trends. To diagnose these trends, we first construct a bucket SST dataset that is corrected for biases common to all groups of bucket SSTs following the same approach used for HadSST3 (refs [4,7]). This reference dataset, ICOADSa, is then further corrected for offsets between groups of ships to obtain our best estimate of SST trends, ICOADSb (see Methods). Comparing the trends in ICOADSa (Fig. 3a) against the difference in trends between ICOADSa and ICOADSb between 1908 and 1941 (Fig. 3b) shows a spatial anticorrelation (Pearson's r-value) of -0.49 ± 0.03. All uncertainties are reported with two standard deviations (s.d.) unless otherwise noted. Similar anticorrelations between our groupwise trend corrections and baseline SST trends are found for each of four major SST products (Table 1), indicating that some of the structure in early-twentieth-century SST trends reflects offsets between nations and decks.

The corrections included in ICOADSb result in a more homogeneous warming pattern (Extended Data Fig. 1f). North Pacific trends change from 0.31 ± 0.03 °C per 34 years in ICOADSa to 0.56 ± 0.11 °C per 34 years in ICOADSb, and North Atlantic trends from 0.85 ± 0.03 °C per 34 years to 0.66 ± 0.11 °C per 34 years (Fig. 4). Results are also temporally more homogeneous, with ICOADSa and other SST estimates indicating a slight cooling between 1920 and 1941 in the North Pacific, but ICOADSb showing a continuous warming trend throughout the early twentieth century (Fig. 4a). Furthermore, whereas ICOADSa indicates that 6% of the 5° grid boxes for which SST trends are computed contain significant cooling ($P < 0.05$), ICOADSb indicates that only 2% of grid boxes contain significant cooling (Fig. 3). Our focus is on the 1908–1941 interval for consistency with a previous study[12], but results are similar if neighbouring starting and ending years are chosen instead (Supplementary Table 2).

The prominent revision to Pacific SST trends relates to a change in offsets identified in the Japanese Kobe Collection, where a mean offset of 0.07 ± 0.12 °C between 1908 and 1930 drops to -0.28 ± 0.13 °C between 1935 and 1941 in the North Pacific (Extended Data Fig. 2c).

One indication of why offsets change comes from considering the distribution of trailing digits in the Kobe Collection records, which are distributed across decks 118, 119 and 762. Before 1932, Kobe Collection records come from deck 762, and the distribution of trailing digits in reported SSTs is consistent with measurements that were originally recorded in whole-degrees Fahrenheit, then converted to Celsius, and finally rounded to tenths-of-a-degree (Extended Data Fig. 3a). Such unit conversion and rounding increases noise but results in negligible systematic offsets[24]. Between 1933 and 1941, however, more than 99% of Kobe Collection records come from deck 118 and are archived in whole-degrees Celsius. Examination of the reference manual for deck 118 reveals that digitization of these observations involved truncation[25] (Extended Data Fig. 3d).

If, before truncation, deck 118 contained the same distribution of trailing digits as found in deck 762, truncation would be expected to introduce a cold offset of −0.46 °C. We find an offset of −0.35 ± 0.07 °C, with the smaller magnitude possibly reflecting the presence of additional offsets between decks. Cold offsets identified in the Kobe Collection during the 1930s are robust to whether or not decks 762 and 118 are treated as distinct groups in our analysis, because these decks are well separated in time and our LME model allows offsets to vary temporally (see Extended Data Figs. 2d and 4 for a sensitivity analysis). Kobe Collection deck 119 is also truncated and found to have a cold offset, but there is little consequence for early-twentieth-century trends because deck 119 spans 1951 to 1961. The major biases identified in decks 118 and 119 are associated with data-management practices and are not addressed in physical models used for correcting bucket temperatures. Furthermore, truncation biases in SSTs could, in principle, be identified by comparing against night-time marine air temperatures, but in the case of decks 118 and 119 this other indicator of surface temperature was also truncated.

North Atlantic warming in ICOADSb is revised downwards by −0.20 ± 0.11 °C between 1908 and 1941, primarily as a consequence of the revision of SSTs from deck 156 and German deck 192 (Fig. 4b and Extended Data Fig. 2b). Deck 156, a group without country information, is biased relatively cold by −0.22 ± 0.12 °C during 1908–1941 in the North Atlantic (Extended Data Fig. 2c). Deck 156 contributes 41% of North Atlantic observations between 1908 and 1912, but this contribution drops to less than 1% during World War I and returns only to 7% thereafter. Deck 156 corrections predominantly warm SST estimates earlier in the twentieth century, decreasing the warming trend by −0.05 ± 0.03 °C per 34 years. Another major correction involves making German deck 192 cooler by 0.33 ± 0.13 °C between 1920 and 1941 (Extended Data Fig. 2c), thereby decreasing estimates of North Atlantic warming by −0.08 ± 0.03 °C per 34 years. A possible clue to the origins of this offset is that German deck 192 shows an increasing portion of SSTs that are reported in whole-degrees Celsius after the 1920s (Extended Data Fig. 3c), although no major modification of the data is indicated by available documentation.

A test of whether our groupwise SST corrections are skilful is made by comparing ICOADSa and ICOADSb against land-based air-temperature records. We compare SSTs from the northwest Pacific and northwest Atlantic against adjacent land temperatures from CRUTEM4 (ref. [26]), because these regions show large corrections near a network of land-temperature stations. The Pearson cross-correlation for interannual temperature variability in the northwest Pacific between 1908 and 1941 increases from 0.67 when using ICOADSa to 0.85 when using ICOADSb. Furthermore, the 1908–1941 trend increases from −0.06 ± 0.04 °C per 34 years in ICOADSa to 0.34 ± 0.11 °C per 34 years in ICOADSb, bringing it into agreement with the 0.33 °C per 34 years trend in CRUTEM4 (Table 1 and Extended Data Fig. 5). Improved agreement for the northwest Atlantic is more modest. Northwest Atlantic correlation increases from 0.65 to 0.70 from ICOADSa to ICOADSb, and trends decrease from 0.92 ± 0.05 °C per 34 years to 0.65 ± 0.13 °C per 34 years, but this is still significantly higher than the CRUTEM4 trend of 0.32 °C per 34 years.

The Pacific Decadal Oscillation[18] (PDO) shifts towards an increasingly positive phase over 1908–1941 (ref. [27]). The magnitude of the PDO trend is smaller in ICOADSb than in ICOADSa (Table 1), but both are consistent with trends found in North Pacific sea level pressure fields[28] (Extended Data Fig. 6).

Equally important to changed SST trends is the greater uncertainty estimated to accompany these trends (Extended Data Fig. 7). The global average SST trend in HadSST3 between 1908 and 1941 has a reported 2 s.d. uncertainty of 0.03 °C per 34 years, whereas accounting for groupwise offsets in ICOADSb reveals an uncertainty of 0.10 °C per 34 years. Larger uncertainties in ICOADSb reflect the fact that averaging repeated observations within a group does not decrease systematic groupwise errors. One implication is that the expected correspondence between observed and simulated trends should be revised downwards. These results have implications for attribution of early-twentieth-century warming and extreme events[5,27,29]. Another implication is that the scope for further improvement of regional temperature estimates associated with better diagnosing and correcting for groupwise biases[9] is greater than previously recognized.

Finally, we briefly explore the implications of our results for model–data mismatches during the early twentieth century. Differences in rates of warming in the North Atlantic and North Pacific reduce from 0.54 ± 0.03 °C per 34 years in ICOADSa to 0.10 ± 0.07 °C per 34 years in ICOADSb. These revised interbasin trend differences are consistent with that of 0.00 ± 0.40 °C per 34 years found in the early-twentieth-century simulations from the Fifth Climate Model Intercomparison Project[30] (CMIP5; Extended Data Fig. 1e). But we note that the global-average rate of SST warming in ICOADSb is 0.56 ± 0.10 °C per 34 years, and that the same domain in the CMIP5 ensemble warms by only 0.19 ± 0.17 °C per 34 years—a discrepancy in warming rates noted previously for other SST estimates[27]. The model–data mismatch in rates of overall warming highlights the importance of continuing to investigate forcing, sensitivity and internal variability of the climate system along with corrections to historical SST estimates.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41586-019-1349-2.

Received: 8 February 2018; Accepted: 17 May 2019;
Published online 17 July 2019.

1. Huang, B. et al. Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *J. Clim.* **30**, 8179–8205 (2017).
2. Hirahara, S., Ishii, M. & Fukuda, Y. Centennial-scale sea surface temperature analysis and its uncertainty. *J. Clim.* **30**, 57–75 (2014).
3. Kennedy, J., Rayner, N., Saunby, M. & Millington, S. Bringing together measurements of sea surface temperature made in situ with retrievals from satellite instruments to create a globally complete analysis for 1850 onwards, HadISST2. In *EGU General Assembly Conference Abstracts* Vol. 15 (2013).
4. Kennedy, J., Rayner, N., Smith, R., Parker, D. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 2. Biases and homogenization. *J. Geophys. Res. D* **116**, D14104 (2011).
5. Delworth, T. L. & Knutson, T. R. Simulation of early 20th century global warming. *Science* **287**, 2246–2250 (2000).
6. Tung, K.-K. & Zhou, J. Using data to attribute episodes of warming and cooling in instrumental records. *Proc. Natl Acad. Sci. USA* **110**, 2058–2063 (2013).
7. Folland, C. & Parker, D. Correction of instrumental biases in historical sea surface temperature data. *Q. J. R. Meteorol. Soc.* **121**, 319–367 (1995).
8. Smith, T. M. & Reynolds, R. W. Bias corrections for historical sea surface temperatures based on marine air temperatures. *J. Clim.* **15**, 73–87 (2002).
9. Kent, E. C. et al. A call for new approaches to quantifying biases in observations of sea surface temperature. *Bull. Am. Meteorol. Soc.* **98**, 1601–1616 (2017).
10. Chan, D. & Huybers, P. Systematic differences in bucket sea surface temperature measurements amongst nations identified using a linear-mixed-effect method. *J. Clim.* **32**, 2569–2589 (2019).
11. Davis, L. L., Thompson, D. W., Kennedy, J. J. & Kent, E. C. The importance of unresolved biases in 20th century sea-surface temperature observations. *Bull. Am. Meteorol. Soc.* **100**, 621–629 (2018).
12. Tokinaga, H., Xie, S.-P. & Mukougawa, H. Early 20th-century Arctic warming intensified by Pacific and Atlantic multidecadal variability. *Proc. Natl Acad. Sci. USA* **114**, 6227–6232 (2017).

13. Crook, J. A. & Forster, P. M. A balance between radiative forcing and climate feedback in the modeled 20th century temperature response. *J. Geophys. Res. D* **116**, D17108 (2011).
14. Knutson, T. R., Zeng, F. & Wittenberg, A. T. Multimodel assessment of regional surface temperature trends: CMIP3 and CMIP5 twentieth-century simulations. *J. Clim.* **26**, 8709–8743 (2013).
15. Stott, P. A. et al. External control of 20th century temperature by natural and anthropogenic forcings. *Science* **290**, 2133–2137 (2000).
16. Myhre, G. et al. Anthropogenic and natural radiative forcing. *Clim. Change* **423**, 658–740 (2013).
17. Schlesinger, M. E. & Ramankutty, N. An oscillation in the global climate system of period 65–70 years. *Nature* **367**, 723–726 (1994).
18. Mantua, N. J., Hare, S. R., Zhang, Y., Wallace, J. M. & Francis, R. C. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bull. Am. Meteorol. Soc.* **78**, 1069–1079 (1997).
19. Karl, T. R. et al. Possible artifacts of data biases in the recent global surface warming hiatus. *Science* **348**, 1469–1472 (2015).
20. Medhaug, I., Stolpe, M. B., Fischer, E. M. & Knutti, R. Reconciling controversies about the 'global warming hiatus'. *Nature* **545**, 41–47 (2017).
21. Thompson, D. W., Kennedy, J. J., Wallace, J. M. & Jones, P. D. A large discontinuity in the mid-twentieth century in observed global-mean surface temperature. *Nature* **453**, 646–649 (2008).
22. Freeman, E. et al. ICOADS release 3.0: a major update to the historical marine climate record. *Int. J. Climatol.* **37**, 2211–2232 (2017).
23. Bonferroni, C. Teoria statistica delle classi e calcolo delle probabilita. *Pubbl. R Ist. Sup. Sci. Econ. Commer. Fir.* **8**, 3–62 (1936).
24. Rhines, A., Tingley, M. P., McKinnon, K. A. & Huybers, P. Decoding the precision of historical temperature observations. *Q. J. R. Meteorol. Soc.* **141**, 2923–2933 (2015).
25. Wilkinson, C. et al. Recovery of logbooks and international marine data: the RECLAIM project. *Int. J. Climatol.* **31**, 968–979 (2011).
26. Jones, P. et al. Hemispheric and large-scale land-surface air temperature variations: an extensive revision and an update to 2010. *J. Geophys. Res. D* **117**, D05127 (2012).
27. Hegerl, G. C., Brönnimann, S., Schurer, A. & Cowan, T. The early 20th century warming: anomalies, causes, and consequences. *Wiley Interdiscip. Rev. Clim. Change* **9**, e522 (2018).
28. Compo, G. P. et al. The twentieth century reanalysis project. *Q. J. R. Meteorol. Soc.* **137**, 1–28 (2011).
29. Seager, R. et al. Would advance knowledge of 1930s SSTs have allowed prediction of the Dust Bowl drought? *J. Clim.* **21**, 3261–3281 (2008).
30. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).

Acknowledgements D.C. and P.H. were funded by the Harvard Global Institute and US National Science Foundation (NSF) Award 1558939. E.C.K. and D.I.B. were funded by the UK Natural Environment Research Council (NERC) through grants NE/R015953/1 and NE/J020788/1. C. Wunsch (MIT, Harvard) and P.-W. Chan (Harvard) provided helpful feedback on an earlier draft.

Author contributions D.C. and P.H. conceived and designed the study; D.C. performed the analyses; and all authors contributed to interpreting results and writing the paper.

Competing interests The authors declare no competing interests.

Additional information
Extended data is available for this paper at https://doi.org/10.1038/s41586-019-1349-2.
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1349-2.
Reprints and permissions information is available at http://www.nature.com/reprints.
Correspondence and requests for materials should be addressed to P.H.
Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

METHODS

Identification of bucket measurements. To identify bucket measurements, we follow the same procedure used for HadSST3 (ref. [4]), using World Meteorological Organization Report Number 47 (WMO47) and ICOADS metadata. Prior to 1941, all SST measurements are assumed to be from buckets unless explicitly recorded otherwise. Analysis of the amplitude of the diurnal cycle in SST before 1941 supports unidentified records as being overwhelmingly from buckets[31]. From 1941 onwards, if the method of measurement is missing in both WMO47 and ICOADS metadata, SST measurements are assumed to come from buckets if the associated nations are reported to have at least 95% of their ships making bucket measurements in WMO47.

The nationality corresponding to each bucket SST measurement is determined using ICOADS country-code information and WMO47. If these metadata are unavailable, the first three letters of the ICOADS identification code are matched with international call signs[32], or the first two letters for decks 705, 706 and 707 (ref. [33]). Decks that may include substantial amounts of measurements not coming from buckets (decks 740, 780 and 874) are discarded, but this has little influence because these decks together contribute only 249 measurements between 1908 and 1941 out of a total of 20.5 million measurements.

Bucket corrections. The same methodology used for HadSST3 (ref. [4]) is applied to ICOADSa and ICOADSb to correct for biases common to all groups (see Extended Data Fig. 8 for comparison and Supplementary Table 3 for details). Corrections common to all groups are made using wooden- and canvas-bucket models[7] run at 5° × 5° for individual climatological months. Bucket models are driven by the 1973–2002 monthly climatology of SST, 10-m air temperature, wind speed and specific humidity from the National Oceanography Centre (NOC) version 2.0 surface flux and meteorological dataset[34] and an insolation climatology from ERA-interim reanalysis[35].

Additional corrections are applied to groups of bucket observations in ICOADSb that are determined using an LME model:

$$\boldsymbol{\delta T} = X\boldsymbol{\alpha} + Z_y\boldsymbol{\beta}_y + Z_r\boldsymbol{\beta}_r + \boldsymbol{\beta}_\sigma \quad (1)$$

The vector of temperature differences, $\boldsymbol{\delta T}$, is determined from proximal pairs of bucket SST observations that come from ships within 300 km and 2 days of one another that are associated with different nationalities and deck assignments. All bucket SST data identified in ICOADS3.0 between 1850 and 2014 are analysed, yielding 17.8 million paired SST differences (Extended Data Fig. 9). The 1908–1941 period contains a subset of 6.1 million SST differences. SST differences are adjusted for climatological effects associated with location, day of year, and hour of day[10].

SST differences contained in $\boldsymbol{\delta T}$ are represented as a 'fixed-effect' term describing offsets between groups, $\boldsymbol{\alpha}$, and random effects describing temporal variations (five-year blocks), $\boldsymbol{\beta}_y$, and regional variations (17 sub-basin regions), $\boldsymbol{\beta}_r$. Matrices X, Z_y and Z_r specify, respectively, common pairs of groups, five-year blocks, and region. $\boldsymbol{\beta}_\sigma$ is the residual, and estimates are derived using an expectation-maximization procedure[36]. Groupwise SST corrections are applied in ICOADSb by removing estimated offsets from each SST measurement according to group, year and region.

Equation (1) is run at two levels, one for determining international offsets and one, a more detailed level, for determining interdeck offsets within nations. Each level of offsets is constrained to equal zero when summed across all paired measurements and all years. The groupwise corrections applied to ICOADSb thus adjust for offsets between groups but do not alter the average across all data. A detailed description of the LME design and implementation, along with the sensitivity of results to plausible variants, is available in a methods paper[10]. In an update to ref. [10], the analysis presented here uses international call signs for identifying nationality, thereby allowing us to increase the number of groups for which more than 5,000 SSTs are compared from 96 to 162 (Extended Data Fig. 10). Only groups associated with at least 5,000 pairs of SST observations are retained.

Trend estimates. Regional trends are the average of local trend estimates resolved at 5° × 5°grid boxes at monthly resolution. Monthly errors are represented as the sum of four different components, $e \approx \mathrm{N}(0, \sigma_o^2 + \sigma_s^2 + \sigma_b^2 + \sigma_g^2)$. Terms represent the uncertainty due to errors associated with individual observations, σ_o; partial sampling of each grid box, σ_s; HadSST3-type bucket-adjustment errors, σ_b; and errors common to individual groups of SST measurements, σ_g. The first three terms are assumed to follow those reported for HadSST3 (ref. [4]), and the last is inferred through the LME model.

A 1,000-member ensemble of SST observations is generated to represent both the random and the systematic components of uncertainty in ICOADSb. In addition to uncertainties that are equivalent to those in HadSST3 (ref. [37]), groupwise bucket errors are included according to the results from our LME model. In particular, intergroup offset terms are drawn from a multivariate normal distribution that represents uncertainties associated with fixed group effects as well as random five-yearly and regional effects[10].

To compute trends, monthly anomalies are averaged to annual values, and years with fewer than three months of data are discarded. Empty monthly 5° grid boxes are infilled by averaging neighbouring grid boxes that are within 10°. Trends are reported only if SST estimates are present (or have been infilled) for both the first and the last five years, and data cover at least 26 of the years between 1908 and 1941. Trends are computed using standard linear least squares. The same procedure is followed for computing trends from the ensemble of realizations in order to estimate uncertainties.

The average trend uncertainty between 1908 and 1941 across sampled grid boxes is 0.12 °C per 34 years for $\sqrt{\sigma_o^2 + \sigma_s^2}$, 0.01 °C per 34 years for σ_b, and 0.06 °C per 34 years for σ_g. When taking spatial averages, σ_o and σ_s are independent across boxes, whereas σ_b is globally systematic and σ_g is partially systematic. The contributions of uncorrelated terms to uncertainties in the global mean trend are essentially negligible for σ_o and σ_s, remain at 0.01 °C per 34 years for σ_b, and become 0.05 °C per 34 years for σ_g. Groupwise errors are thus expected to dominate the uncertainties associated with large-scale SST trends.

Comparison with other datasets. There are a number of notable differences between the SST datasets considered here. ICOADSa, ICOADSb, and ERSST5 (ref. [1]) are based on ICOADS3.0 (ref. [22]), whereas COBESST2 (ref. [2]), HadISST2 (ref. [3]) and HadSST3 (ref. [4]) are based on ICOADS2.5 (ref. [38]). ICOADSa and ICOADSb use only bucket SST measurements, which are estimated to account for 94% of all observations in ICOADS3.0 between 1908 and 1941. Other datasets also make use of engine-room intake, buoy and drifter observations, which become more common after 1941. Finally, ERSST5, COBESST2 and HadISST2 infill monthly grid boxes without data, whereas ICOADSa, ICOADSb and HadSST3 leave these boxes unfilled.

Coastal near-surface air temperatures from CRUTEM4 (ref. [26]) are used to check the validity of our groupwise corrections. We choose to compare ICOADSa and ICOADSb with CRUTEM4 near the east coasts of Asia and North America because these two regions experience the largest adjustments in trends and because of the availability of relatively dense station and bucket data. In each area, regional SST and land air temperature time series are computed using only those grid boxes that contain both types of measurement (Extended Data Fig. 5c).

Pacific Decadal Oscillation. An SST index for the Pacific Decadal Oscillation, SST-PDO, is obtained by projecting annual-average SSTs poleward of 20° N in the Pacific onto a normalized PDO pattern, where the PDO pattern is obtained by regressing SST onto a standard National Centers for Environmental Information (NCEI) PDO index[18] over the years 1948–2014 and is then normalized to have zero mean and a range of one over the North Pacific. The same method is applied to sea level pressure (SLP) using the NOAA 20th Century Reanalysis[28], yielding SLP-PDO. SST-PDO is regressed against SLP-PDO over the periods 1908–1941 and 1948–2010 for sensitivities in units of °C mb^{-1}. Uncertainty in PDO trends is estimated by randomly perturbing PDO indices using error estimates of projections in individual years. When estimating the sensitivity of the SST-PDO index against the SLP-PDO index, the same random seeding is used to draw realizations of both indices for all SST products.

Data availability

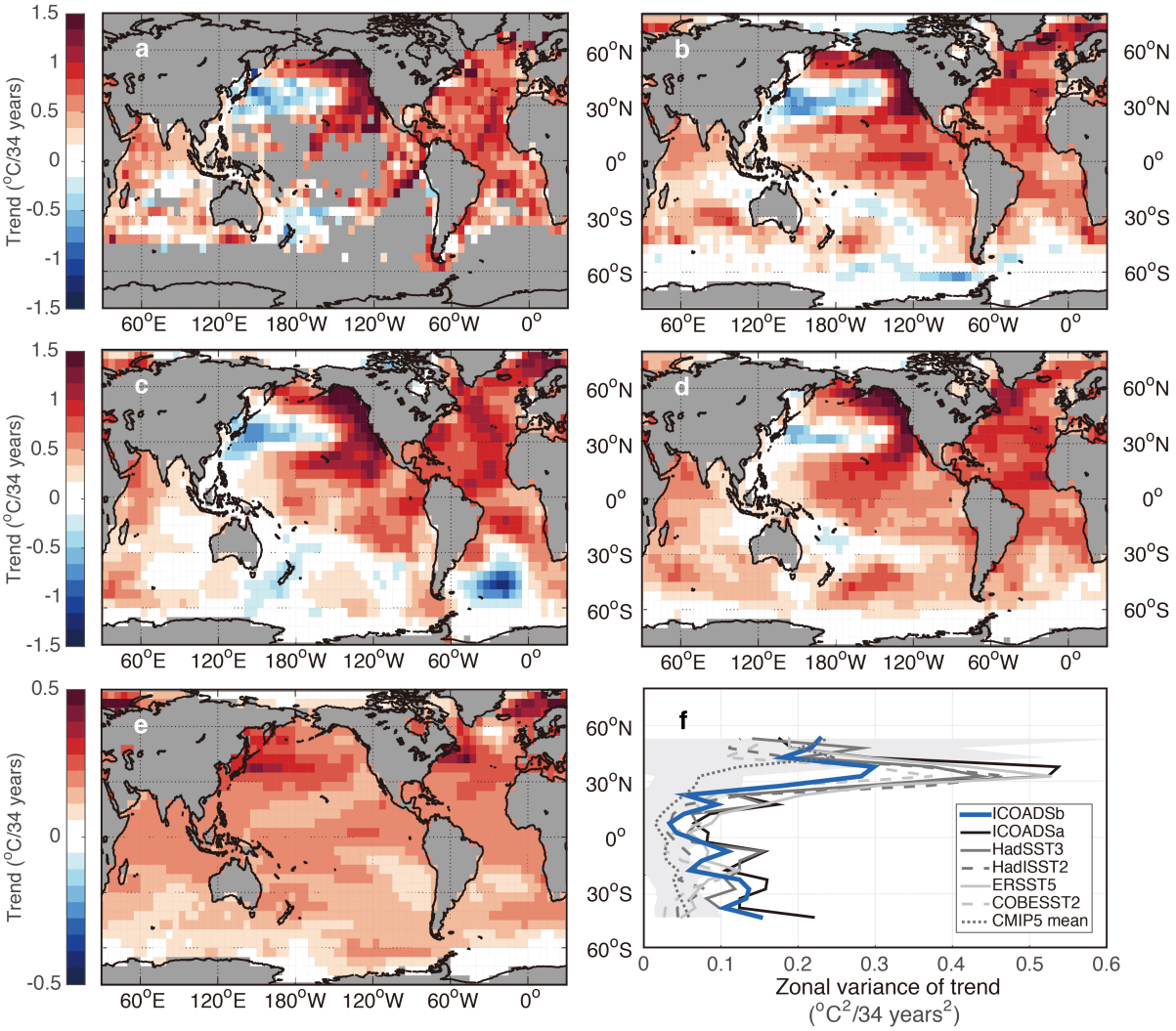
All datasets used in this study are publicly available as follows: ICOADS3.0 (https://rda.ucar.edu/datasets/ds548.0/), HadSST3 and a 100-member ensemble (https://www.metoffice.gov.uk/hadobs/hadsst3/data/download.html), HadISST2 and a 10-member ensemble (https://www.metoffice.gov.uk/hadobs/hadisst/data/hadisst2/), COBESST2 (https://www.esrl.noaa.gov/psd/data/gridded/data.cobe2.html), ERSST5 (https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.ersst.v5.html), CRUTEM4 (https://crudata.uea.ac.uk/cru/data/temperature/), ERA-Interim Reanalysis (https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era-interim), NOAA 20th Century Reanalysis (https://www.esrl.noaa.gov/psd/data/20thC_Rean/), and NOCv2.0 surface flux (https://rda.ucar.edu/datasets/ds260.3/). ICOADSa and ICOADSb, together with archived versions of all other SST datasets used in this study, are posted at https://doi.org/10.7910/DVN/DXJIGA.

Code availability

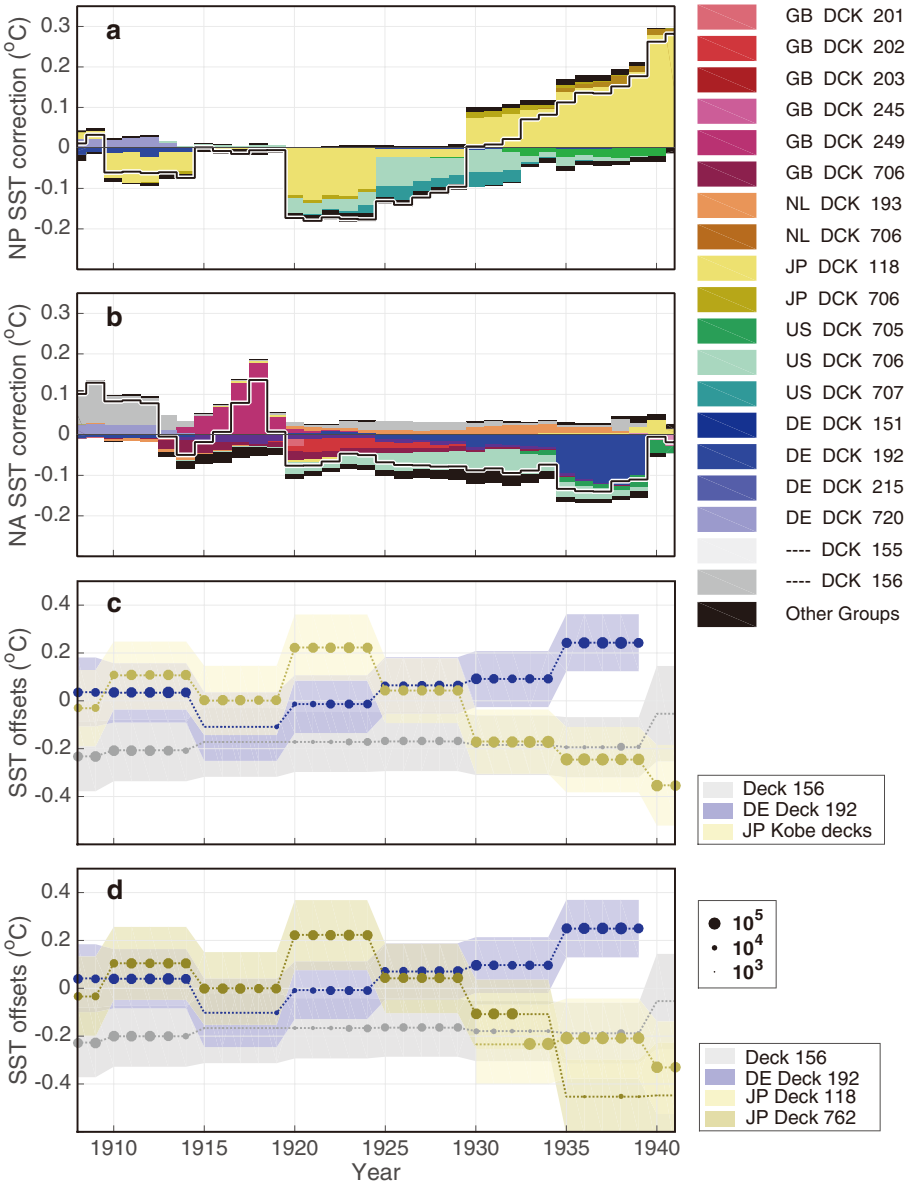
Code allowing the full reproduction of our results is posted on Github at https://github.com/duochanatharvard/Homogeneous_early_20th_century_warming.

31. Carella, G. et al. Estimating sea surface temperature measurement methods using characteristic differences in the diurnal cycle. *Geophys. Res. Lett.* **45**, 363–371 (2018).
32. International Telecommunications Union *Radio Regulations* Appendix 2, 793–799 (ITU, 2016).
33. Carella, G., Kent, E. C. & Berry, D. I. A probabilistic approach to ship voyage reconstruction in ICOADS. *Int. J. Climatol.* **37**, 2233–2247 (2017).
34. Berry, D. I. & Kent, E. C. A new air–sea interaction gridded dataset from ICOADS with uncertainty estimates. *Bull. Am. Meteorol. Soc.* **90**, 645–656 (2009).

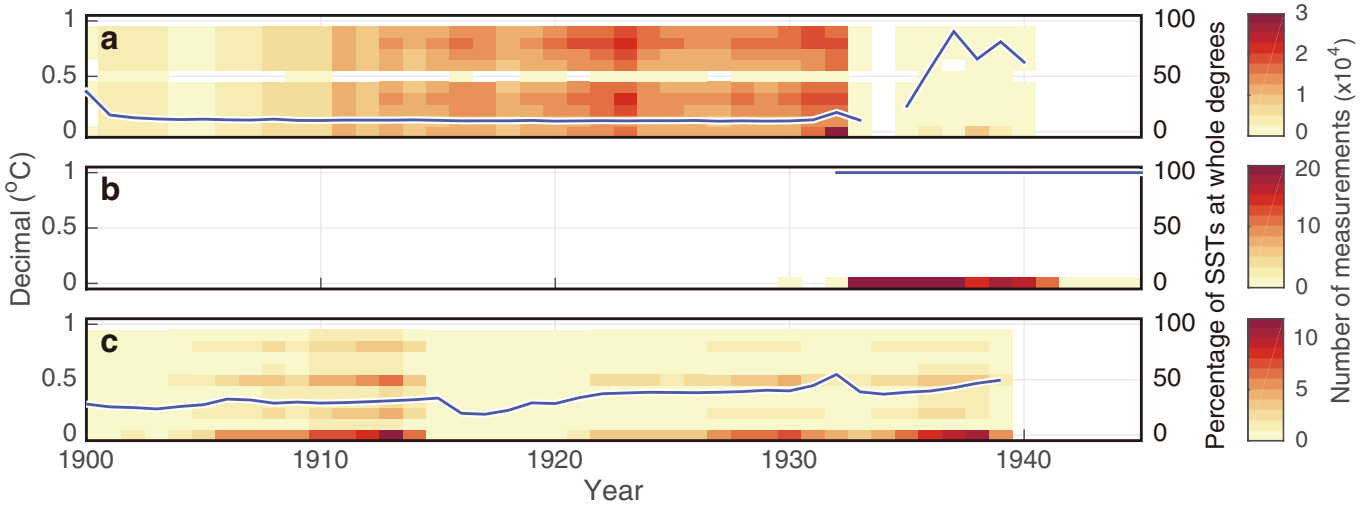
35. Dee, D. P. et al. The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
36. Harville, D. A. Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Stat. Assoc.* **72**, 320–338 (1977).
37. Kennedy, J., Rayner, N., Smith, R., Parker, D. & Saunby, M. Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *J. Geophys. Res. D* **116**, D14104 (2011).
38. Woodruff, S. D. et al. ICOADS release 2.5: extensions and enhancements to the surface marine meteorological archive. *Int. J. Climatol.* **31**, 951–967 (2011).



Extended Data Fig. 1 | 1908–1941 SST trends in major observational estimates and CMIP5 simulations. a–d, Maps of SST trends for: **a**, HadSST3 (ref. [4]); **b**, HadISST2 (ref. [3]); **c**, ERSST5 (ref. [1]); and **d**, COBESST2 (ref. [2]). Although different correction schemes were used to account for bucket biases in each observational estimate, panels **a–d** show similar early-twentieth-century trends. **e**, Averages from 88 CMIP5 historical experiment simulations[30]. Model results and observed trends are all regridded to a common 5°×5° resolution to facilitate intercomparison. **f**, Zonal variance in 1908–1941 SST trends. Shading indicates a range covering 80% of the 88 CMIP5 members, which come from 39 coupled climate models (see Supplementary Table 4). ICOADSb has a more uniform spatial pattern of early-twentieth-century warming than other SST estimates.



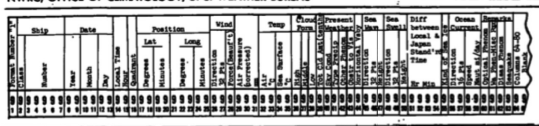
Extended Data Fig. 2 | Groupwise decomposition of SST corrections in ICOADSb. a, **b**, Contributions from individual groups over the North Pacific Ocean (NP; **a**) and the North Atlantic Ocean (NA; **b**). **c**, SST offsets for groups having major influences on 1908–1941 trend estimates. Groupwise corrections relate foremost to a growing cold bias amongst Japanese deck 118 (yellow) in the Pacific, a pre-World War I cold bias in deck 156 (grey), and a growing warm bias in German deck 192 (blue) in the Atlantic. Shading indicates 2 s.d. uncertainties, and the sizes of the markers indicate numbers of SST observations. **d**, As for **c**, but for an analysis in which the Japanese Kobe collection is divided into decks 762, 118 and 119. (Deck 119 is not shown because it spans only 1951 to 1961.) In this analysis, the mean offset over the North Pacific is 0.07 ± 0.14 °C (2 s.d.) for deck 762 from 1908 to 1930, and -0.24 ± 0.16 °C for deck 118 from 1935 onward, consistent with the case when decks are not separated (**c**). In addition, the cold offset of the Japanese Kobe Collection and the warm offset of German deck 192 are robust even if SSTs from the two groups are not allowed to be directly paired in the intercomparison.



d

CLIMATIC CENTER, USAF, AIR WEATHER SERVICE (MATS)
NWRC, OFFICE OF CLIMATOLOGY, U. S. WEATHER BUREAU

REFERENCE MANUAL 118 JAPANESE SURFACE MARINE OBS.



CARD CONTENT

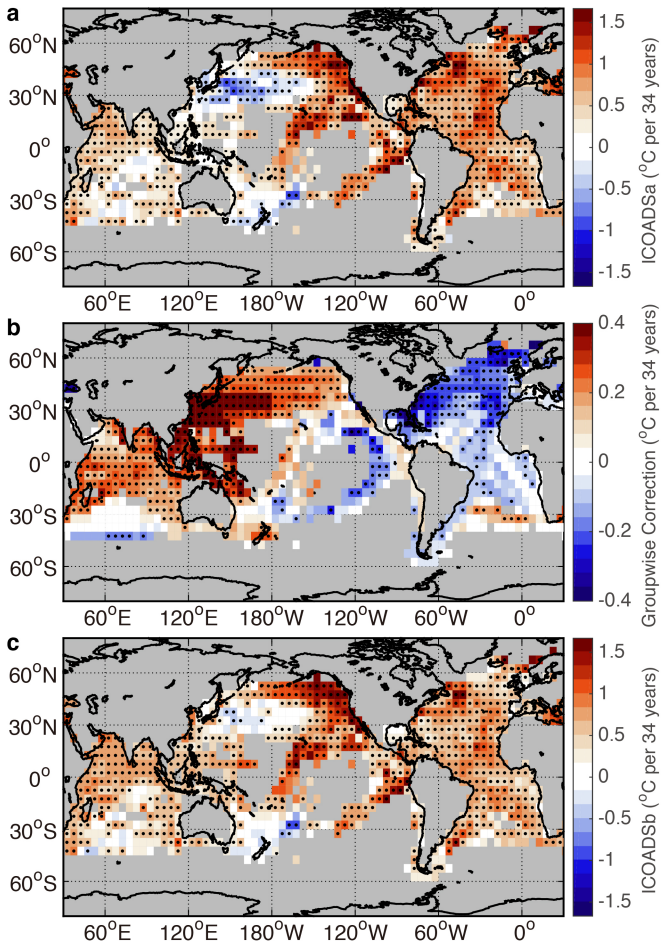
Column	Item or Element	Symbolic Letter	Card Code	Card Code Definition	Remarks
1-63	All elements		Blank	Missing Data	Card columns left blank in columns 1-63 indicate missing data unless otherwise indicated in the respective codes. All elements were converted to the described codes and units before punching.
1	Format		1	Format Number for Deck 118	
2	Ship class		0-5	See Code 1.	
3-7	Ship Number		00000-99999	Indicates ship's name	Arbitrarily assigned number for each ship stamped on the upper left side of record.
8-9	Year		37-53	1933-1953	
10-11	Month		01-12	January-December	
12-13	Day of Month		01-31		
14-15	Hour LMT		06, 12, 18, 24		See Page 1, Observation Time.
16	Quadrant		0-3,	See Code 2.	
17-20	Latitude				North and South indicated in column 16
17-18	Degrees	*	00-90	degrees and	
19-20	Minutes	*	00-59	minutes	
21-25	Longitude				East and West indicated in column 16.
21-23	Degrees	*	000-180	degrees and	
24-25	Minutes	*	00-59	minutes	
26-27	Wind Direction		00-32	See Code 3.	The 16 points of a 32 point compass is most frequently used.
28	Wind Force	F	0-9 X/0 X/1 X/2	Beaufort Force 0-9 Beaufort Force 10 Beaufort Force 11 Beaufort Force 5 12 See Code 4	Hurricane 5 73 m.p.h. or 64 knots
29-31	Barometric Sea Level Pressure (Corrected)		000-999 X/000-X/999	700.0 through 799.9mm in 1/10 millimeters 600.0 through 699.9mm	At standard height 6m. above water X-overpunch in column 29 indicates the hundreds position value is 6 rather than 7.

CARD CONTENT

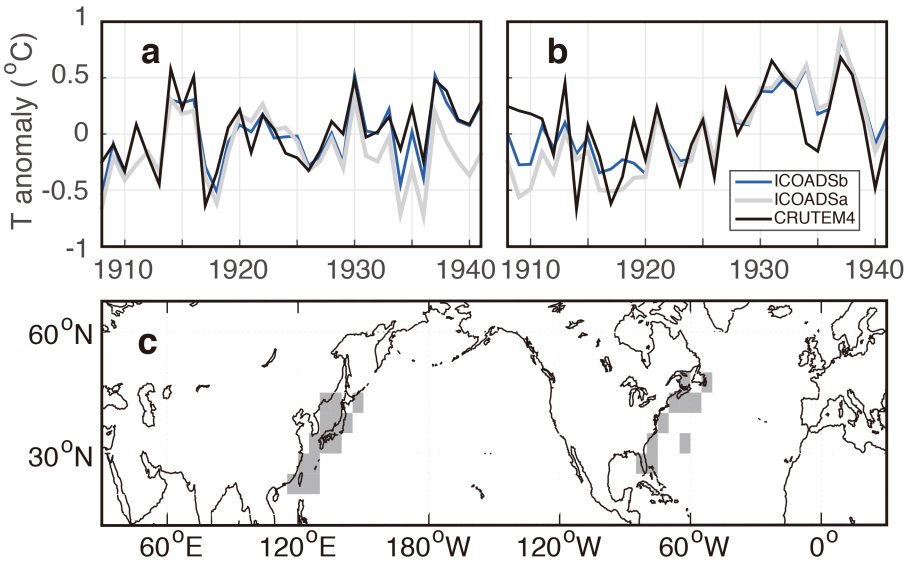
Column	Item or Element	Symbolic Letter	Card Code	Card Code Definition	Remarks
32-33	Air Temperature	TT	00-99 X/01-X/99	0 through 99 °C -1 through -99 °C	1/10 °C values are dropped and punched in whole °C. X-overpunch in tens position equals minus (-) values.
34-35	Sea Surface Temperature		00-99 X/01-X/99	0 through 99 °C -1 through -99 °C	1/10 °C values were dropped and punched in whole °C. X-overpunch in tens position equals minus (-) values.
35-39	Clouds				
36	High Type	C_H	0-7	See Code 5.	
37	Middle Type	C_M	0-3	See Code 6.	
38	Low Type	C_L	0-9	See Code 7.	
39	Total Amount		0-9 X	0-9/10 sky covered 10/10, total sky covered or obscured	
40-43	Present Weather				
40	Sky Condition		Blank 0-5	See Code 8.	
41	Type of Precipitation		Blank 0-9 X/0-X/5	See Code 9.	
42	Other Phenomena		Blank 0-9, X	See Code 10.	
43	Obstruction to Vision		Blank 0-3	See Code 11.	
44	Visibility	V	0-9	See Code 12.	
45-46	Direction of Sea Waves		00-32, XX	See Code 3.	See Remarks Columns 26-27.
47	Sea Waves Height	S	0-9, X	See Code 13.	
48-49	Direction of Sea Swell		00-32	See Code 3.	See Remarks Columns 26-27.
50	Swell Height	K	0-7	See Code 14.	See page 1, Additional Remarks, paragraph 2.
51-54	Time Difference between ships and JST		00-21	Ship's time is 00-21 hours slow from JST	See page 1, Observation Time. X in column 51 and Blank in columns 52-54 indicate time difference missing
51-52	Hours		X0-X3 or 30-33	Ship's time is 0-3 hours fast from JST	
53-54	Minutes		00-59	00-59 Minutes	
55	Kind of Ice	C_2	0-9, X	See Code 15.	Covers the preceeding 6 hours.
56-57	Current Direction		00-36	See Code 16.	Direction toward which ocean current is moving Normally reported once daily on the 1200 LMT.

Issue Date: 18 Apr 61 Page 2

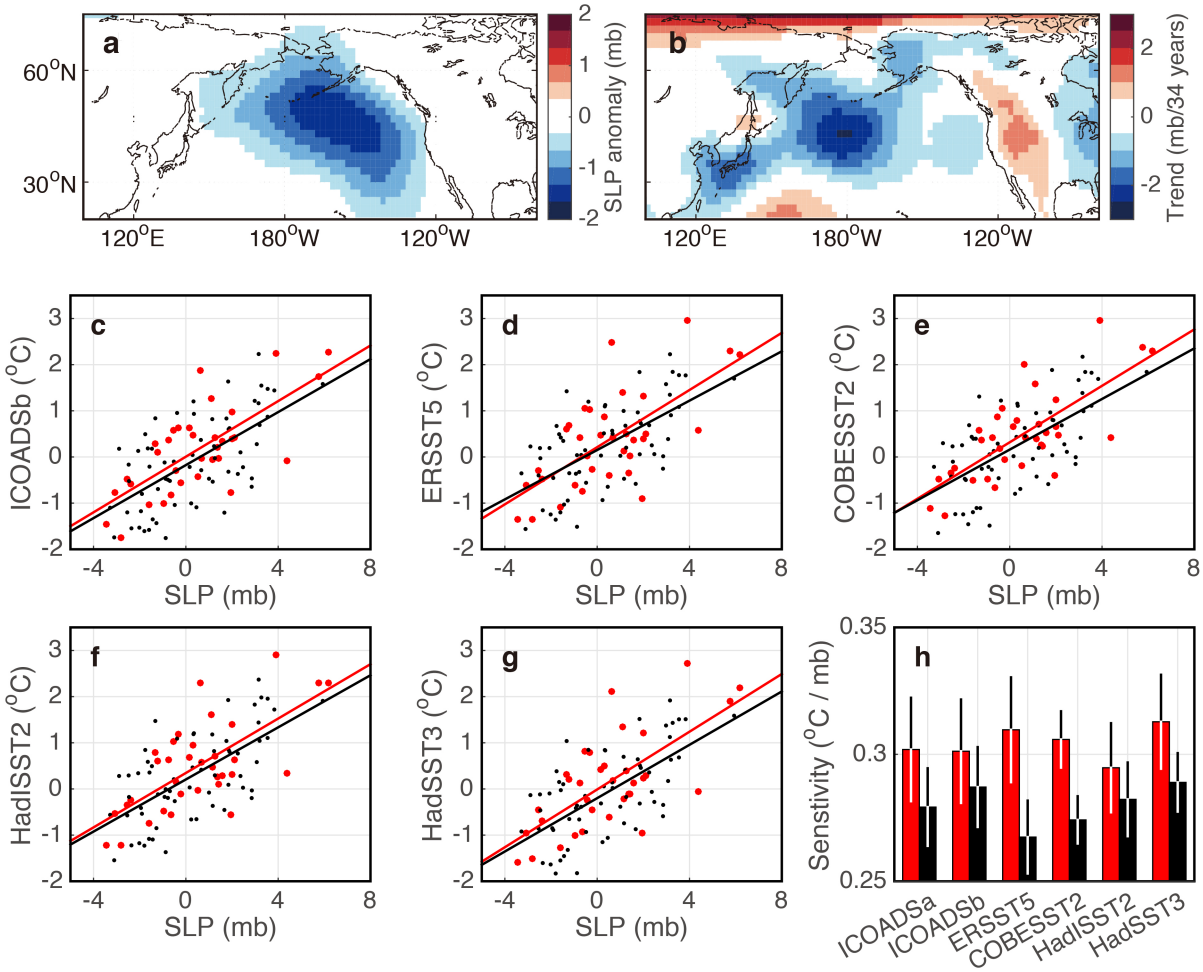
Extended Data Fig. 3 | Decimal distributions of SSTs. a, Deck 762 from the Japanese Kobe Collection. The near-absence of a 0.5 °C decimal is indicative of data being recorded in whole-degrees Fahrenheit and then converted into tenths-of-a-degree Celsius. The blue line shows the percentage of SSTs recorded in whole degrees. **b**, Temperatures reported in deck 118 of the Japanese Kobe Collection are all truncated to whole-degrees Celsius and constitute more than 99% of the Kobe Collection measurements between 1933 and 1941. **c**, German deck 192 is a time-varying mix of data in whole-degrees Celsius and data in tenths-of-a-degree Celsius; most common in the latter are values of 0 and 5, and then even digits. The percentage of SSTs archived in whole-degrees Celsius increases from 29% in 1908 to 50% in 1941 (blue line). **d**, Image from the reference manual of Japanese Kobe Collection deck 118 (ref. [25]), where the red box highlights that temperature measurements were digitized at whole-degrees Celsius and all decimals dropped (https://icoads.noaa.gov/reclaim/pdf/dck118.pdf).



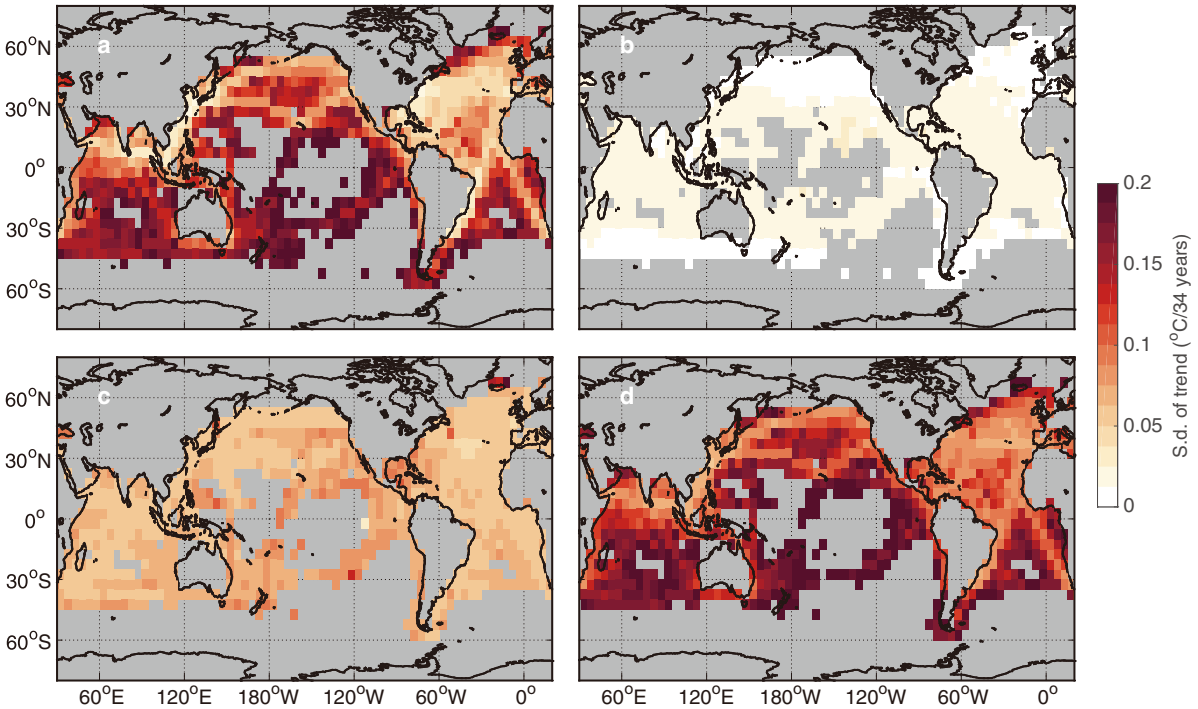
Extended Data Fig. 4 | Sensitivity of 1908–1941 SST trends. As for Fig. 3, but with Japanese Kobe Collection decks 118, 119 and 762 separated into distinct groups. **a**, SST trends in ICOADSa; **b**, trends associated with the corrections for groupwise offsets; **c**, SST trends in ICOADSb after applying groupwise corrections.



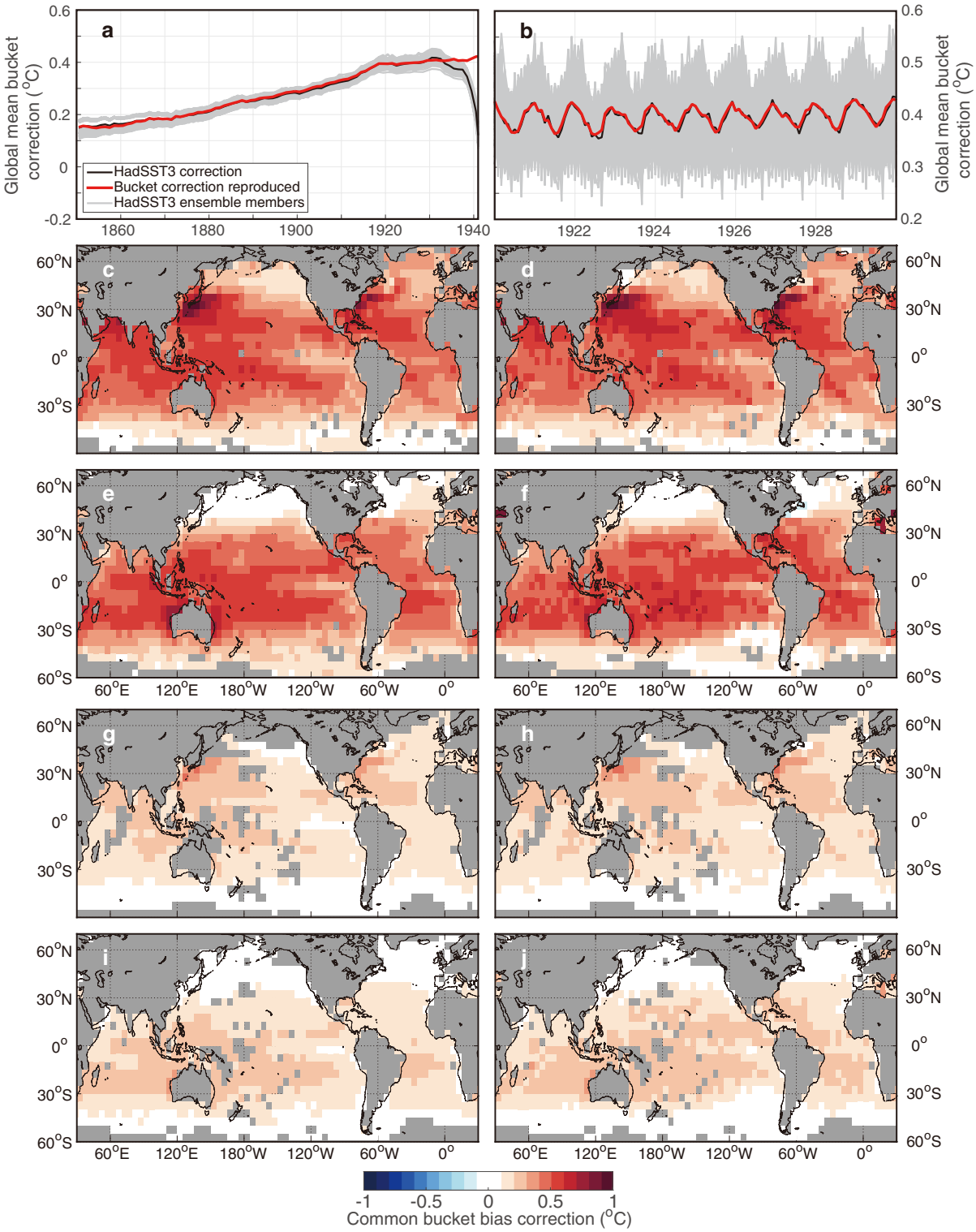
Extended Data Fig. 5 | Comparison of SSTs with coastal air temperature estimates. a, **b**, Air temperatures are from CRUTEM4 (ref. [26]) near the east coast of Asia (**a**) and the east coast of the US (**b**). These two regions are indicated by the shading in **c** and host the largest groupwise corrections. SSTs corrected for groupwise offsets (ICOADSb, blue line) correlate better with CRUTEM4 air temperatures (black line) than do SSTs having only common bucket-bias corrections (ICOADSa, grey line), especially for Asia in the 1930s. See Table 1 for correlations associated with other major SST estimates.



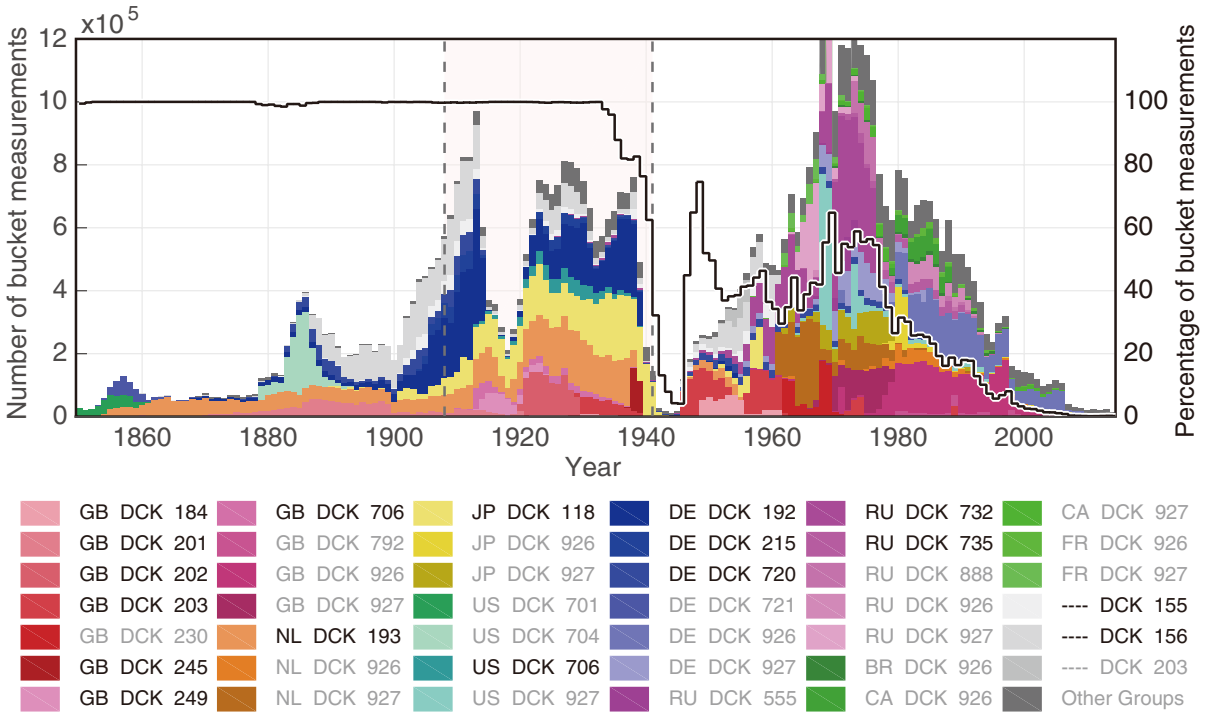
Extended Data Fig. 6 | Sea level pressure and the Pacific Decadal Oscillation. a, Spatial pattern of sea level pressure (SLP) associated with a 1σ increase in the PDO index[18]. **b**, 1908–1941 SLP trends from the NOAA's 20th Century Reanalysis[28]. **c**–**g**, Regressions of SST-PDO indices against the SLP-PDO index for ICOADSb (**c**), ERSST5 (**d**), COBESST2 (**e**), HadISST2 (**f**) and HadSST3 (**g**). The SLP-PDO index is computed from the NOAA's 20th Century Reanalysis, and the regression analysis is performed for 1908–1941 (red) and 1948–2010 (black). **h**, Regression coefficients for each dataset over 1908–1941 (red) and 1948–2010 (black). Note that the *y* axis starts from 0.25 °C mb^{-1}. Error bars indicate 2 s.d. uncertainties. ICOADSb has a similar regression coefficient over 1908–1941 to that found using other SST datasets and other epochs.



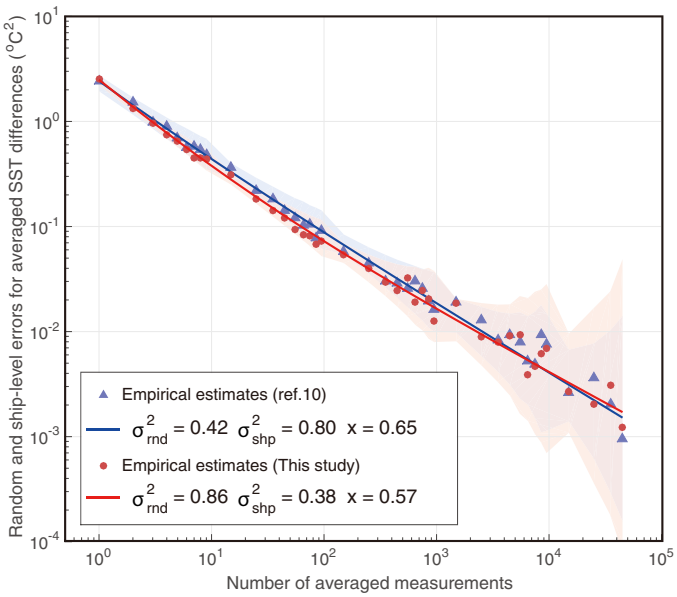
Extended Data Fig. 7 | 1σ uncertainty estimates of 1908–1941 trends. **a**, Sampling and observational uncertainty inherent to all SST estimates, estimated using 1,000 members drawn from normal distributions that have a mean of zero, and a variance of the sampling and observational uncertainty equal to that specified in ref. [37]. **b**, Uncertainties associated with common bucket-bias corrections in HadSST3, which takes into account shifts between wooden and canvas buckets, exposure time, and ship speed, obtained from a 100-member ensemble[4]. **c**, Uncertainty of groupwise corrections in ICOADSb, estimated from 1,000 random correction members. **d**, Uncertainties quadratically combining those shown in **a**–**c**. Note that these uncertainties are for individual grid boxes, whereas Table 1 reports uncertainties for regional or global averages.



Extended Data Fig. 8 | Common bucket-bias corrections. a, Comparison of annual global mean common bucket-bias corrections reproduced in ICOADSa and from HadSST3 for 1850–1941. **b**, As in **a**, but for monthly mean corrections from 1920 to 1929. Bucket corrections in ICOADSa are subsampled over HadSST3 grids before computing global means. **c**–**f**, Comparisons of spatial patterns of SST corrections for the period 1920 to 1930. Shown are our corrections for December, January and February (**c**) and June, July and August (**e**); and the HadSST3 corrections for December, January and February (**d**) and June, July and August (**f**). The 1920–1930 period is assumed to contain only canvas buckets, and an average of 78% of SSTs from fast ships (7 m s^{-1}) and 22% from slow ships (4 m s^{-1}). **g**–**j**, As for **c**–**f**, but for corrections over 1850 to 1860—a period assumed to contain only slow ships (4 m s^{-1}), with 36% of SST measurements from canvas buckets and 64% from wooden buckets. Discrepancies become apparent from the 1930s because of increasing contributions from warm-biased engine-room-intake measurements in HadSST3, whereas we focus only on bucket measurements.



Extended Data Fig. 9 | Groups providing bucket SST measurements from 1850 to 2014. Left-hand *y*-axis: numbers of bucket SST measurements from individual groups, identified by country and deck information as colour coded, in ICOADS3.0. CA, Canada; BR, Brazil; DE, Germany; FR, France; GB, Great Britain; NL, Netherlands; JP, Japan; RU, Russia; US, United States; —, missing country information. Black text shows the 46 of 162 groups that contributed to the SST archive between 1908 and 1941, which is indicated by shading. Groups with fewer than 2×10^5 measurements throughout 1850 to 2014 are labelled as 'other groups'. Right-hand *y*-axis, percentage of measurements that are from buckets (black line). Bucket measurements constitute nearly all observations before the 1930s.



Extended Data Fig. 10 | Partitioning of random and systematic errors. Updating the methodology for identifying nationality allows more groups of SSTs to be compared, and leads to minor changes in estimated random and systematic errors. Shown are fits according to equation (6) in ref. [10], depending on random measurement error (σ^2_{rnd}), systematic ship-level error (σ^2_{shp}), and the scaling relationship between the number of measurements and ships (x). Fits are performed using logarithms, and shading indicates 2 s.d. in logarithmic space.

LETTER

https://doi.org/10.1038/s41586-019-1383-0

Notum produced by Paneth cells attenuates regeneration of aged intestinal epithelium

Nalle Pentinmikko[1], Sharif Iqbal[1,14], Miyeko Mana[2,14], Simon Andersson[1], Armand B. Cognetta III[3], Radu M. Suciu[3], Jatin Roper[4], Kalle Luopajärvi[1], Eino Markelin[1], Swetha Gopalakrishnan[1], Olli-Pekka Smolander[1], Santiago Naranjo[2], Tuure Saarinen[5,6], Anne Juuti[6], Kirsi Pietiläinen[5], Petri Auvinen[1], Ari Ristimäki[7], Nitin Gupta[8], Tuomas Tammela[9], Tyler Jacks[2,10], David M. Sabatini[2,10,11], Benjamin F. Cravatt[3], Ömer H. Yilmaz[2] & Pekka Katajisto[1,12,13]*

A decline in stem cell function impairs tissue regeneration during ageing, but the role of the stem-cell-supporting niche in ageing is not well understood. The small intestine is maintained by actively cycling intestinal stem cells that are regulated by the Paneth cell niche[1,2]. Here we show that the regenerative potential of human and mouse intestinal epithelium diminishes with age owing to defects in both stem cells and their niche. The functional decline was caused by a decrease in stemness-maintaining Wnt signalling due to production of Notum, an extracellular Wnt inhibitor, in aged Paneth cells. Mechanistically, high activity of mammalian target of rapamycin complex 1 (mTORC1) in aged Paneth cells inhibits activity of peroxisome proliferator activated receptor α (PPAR-α)[3], and lowered PPAR-α activity increased Notum expression. Genetic targeting of *Notum* or Wnt supplementation restored function of aged intestinal organoids. Moreover, pharmacological inhibition of Notum in mice enhanced the regenerative capacity of aged stem cells and promoted recovery from chemotherapy-induced damage. Our results reveal a role of the stem cell niche in ageing and demonstrate that targeting of Notum can promote regeneration of aged tissues.

Tissue turnover and regenerative capacity decrease with ageing in many tissue types[4–6]. The intestinal epithelium is one of the fastest renewing tissues in the human body and has previously been reported to regenerate without loss of self-renewal in long-term in vitro organoid culture[7]. However, complications in the gastrointestinal system increase with age[8–10], and intestines of old mice regenerate more slowly after radiation-induced damage[11], which suggests reduced stem cell activity.

To assess possible ageing-induced changes in the human intestinal epithelium, we used the capacity of intestinal-stem-cell (ISC)-containing epithelial crypts to form clonogenic organoids[7] as an in vitro assay of intestinal regenerative potential. We observed a significant age-induced reduction in the organoid-forming capacity of colonic crypts derived from biopsies from healthy human donors (Fig. 1a). As the heterogeneous human colon material does not allow downstream analysis of stem-cell-intrinsic and -extrinsic effects, we next analysed the effects of age on mouse small-intestinal epithelium. Crypts from old (more than 24-month-old) mice formed significantly fewer organoids than those isolated from young (3-to-9-month-old) mice (Extended Data Fig. 1a). Notably, regenerative growth of de novo crypts was also diminished in the organoids formed by the old crypts (Fig. 1b, Extended Data Fig. 1b, c), which indicates a reduction in stem cell function. Furthermore, the reduced crypt formation observed during serial passage of secondary crypt domains demonstrated that the decline in epithelial regeneration was due to alterations intrinsic to the epithelium (Extended Data Fig. 1d, e).

Intestinal tissue renewal is largely maintained by the LGR5-expressing ISCs, which are located between Paneth cells at the crypt base. ISCs divide regularly and produce transit-amplifying progenitor cells that divide several additional times and gradually differentiate. Paneth cells produce antimicrobial peptides and multiple signalling factors, such as epidermal growth factor (EGF), Wnt3, Delta-like ligands and cyclic ADP ribose (cADPR)[2,12], which regulate stemness and function of the neighbouring ISCs. To more specifically address the separate roles of stem cells and their niche in age-associated intestinal decline, we used *Lgr5-eGFP-IRES-creERT2* reporter mice[1], which enable identification and isolation of Paneth cells, Lgr5–eGFP^{hi} ISCs, and transit-amplifying cells that can be further divided into immediate eGFP^{med} and late eGFP^{lo} progenitors.

The aged mouse crypts did not present gross histological alterations, and the fraction of ISCs and transit-amplifying cells that were 5-ethynyl-2′-deoxyuridine (EdU)$^{+}$ or Ki67^{+} was unchanged in old mouse and human samples (Extended Data Fig. 1f–h). However, flow cytometry of crypts from old mice revealed a significant drop in frequency of Lgr5^{hi} ISCs (Fig. 1c), whereas Paneth cell frequency was significantly increased in old mice and humans (Fig. 1c, Extended Data Fig. 1i, j). As the *Lgr5-eGFP-IRES-creERT2* mouse model exhibits mosaic expression of the eGFP-containing construct[1], the alterations in cellular frequencies were also validated by immunohistochemical analyses of olfactomedin 4 and lysozyme as markers of ISCs and Paneth cells, respectively (Extended Data Fig. 1k). The reduction in ISC number, together with the unchanged EdU^{+} frequency suggested that old crypts may have a lower output of cells, which possibly contributes to villus blunting and slower intestinal turnover during ageing[13]. As Paneth cells positively regulate the number and function of Lgr5^{hi} stem cells in young mice[2,12], the decoupling of the Lgr5^{hi}:Paneth cell ratio in old mice (Extended Data Fig. 1l) raised the possibility that interactions between these two cell types change during ageing. To address this, we investigated the organoid-forming capacity of co-cultured Lgr5^{hi} and Paneth cells isolated from young and old mice (Fig. 1d). Old Paneth and Lgr5^{hi} cells both showed cell-type-specific age-induced effects, and initiated organoids with reduced efficiency. Consistent with previous work, neither cell type formed organoids efficiently by themselves[2,12] but, when co-cultured with Paneth cells, Lgr5^{hi} cells from young mice formed organoids at a higher rate than old Lgr5^{hi} cells. Of note, the age-induced stem cell defect was partially rescued by co-culture with young Paneth cells, whereas old Paneth cells did not fully support

[1]Institute of Biotechnology, HiLIFE, University of Helsinki, Helsinki, Finland. [2]The David H. Koch Institute for Integrative Cancer Research at MIT, Department of Biology, MIT, Cambridge, MA, USA. [3]The Skaggs Institute for Chemical Biology, Department of Chemical Physiology, The Scripps Research Institute, La Jolla, CA, USA. [4]Department of Medicine, Division of Gastroenterology, Duke University, Durham, NC, USA. [5]Obesity Research Unit, Research Programs Unit, Diabetes and Obesity, University of Helsinki, Helsinki, Finland. [6]Abdominal Center, Department of Gastrointestinal Surgery, Helsinki University Hospital, Helsinki, Finland. [7]Department of Pathology, Research Programs Unit and HUSLAB, University of Helsinki and Helsinki University Hospital, Helsinki, Finland. [8]Atlanta Gastroenterology Associates, Atlanta, GA, USA. [9]Cancer Biology and Genetics, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [10]Howard Hughes Medical Institute, MIT, Cambridge, MA, USA. [11]Whitehead Institute for Biomedical Research, Howard Hughes Medical Institute, Department of Biology, MIT, Cambridge, MA, USA. [12]Molecular and Integrative Bioscience Research Programme, Faculty of Biological and Environmental Sciences, University of Helsinki, Helsinki, Finland. [13]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. [14]These authors contributed equally: Sharif Iqbal, Miyeko Mana. *e-mail: pekka.katajisto@helsinki.fi

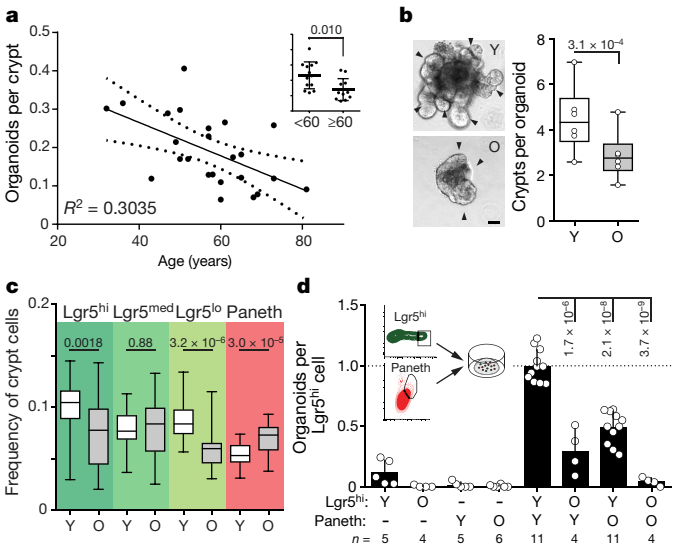


Fig. 1 | Age-associated reduction in intestinal regeneration is caused by decreased function of Lgr5$^+$ stem cells and the Paneth cell niche. **a**, Organoid-forming capacity of human colonic crypts ($n = 24$). Solid and dotted lines show linear regression and 95% confidence interval. Inset, crypts from donors of more than 60 years of age show significantly lower organoid-forming capacity. **b**, Regenerative growth of crypts from young (Y) and old (O) mice. Organoids derived from young mice ($n = 6$) generate more de novo crypt domains (arrowheads) in primary cultures (5–9 days after isolation). Representative images from day seven of culture. Scale bar, 50 μm. Student's paired *t*-test. **c**, Cellular frequencies analysed by flow cytometry ($n = 30$ young, $n = 26$ old). For fluorescence-activated cell sorting (FACS) gating strategy, see Supplementary Fig. 1. **d**, Clonogenicity of young and old Lgr5^{hi} stem cells co-cultured with young and old Paneth cells. *n*, number of mice analysed. *P* values shown for comparison of average for young Lgr5^{hi} cells co-cultured with young Paneth cells. Young mice aged between 3 and 9 months; old mice over 24 months of age in all experiments. Unless otherwise mentioned, in box plots, the line represents median, the box shows interquartile range and whiskers show the range. All other data are represented as mean ± s.d. and conditions compared by two-tailed unpaired Student's *t*-test. *P* values shown in corresponding panels. $P < 0.05$ is considered significant.

organoid formation by young Lgr5^{hi} cells. These data indicate that both stem-cell-intrinsic and -extrinsic epithelial factors reduce the regenerative potential during intestinal ageing.

Decline in fatty acid oxidation was recently shown to intrinsically reduce the function of aged intestinal stem cells[14]. Old Paneth cells extrinsically decreased clonogenic growth of young Lgr5^{hi} cells even in a long-term co-culture (Extended Data Fig. 1m, n). While the original Paneth cells lasted at least 14 days in such co-cultures (Extended Data Fig. 1o), new Paneth cells were continuously produced by the stem cells, which suggests that exposure to old Paneth cells had long-term effects on ISCs and their progeny. To understand the mechanics of how age-induced changes in niche–stem cell communication may influence stem cells, we performed RNA sequencing on both cell types (Supplementary Table 1). Old Paneth cells showed specific deregulation of genes that encode secreted or plasma-membrane-associated proteins (Extended Data Fig. 2a, b). Among the key stemness-regulating factors, we noted no alterations in expression of *Wnt3* or *Egf*, whereas expression of cADPR-producing *Bst1* was reduced (Extended Data Fig. 2c). However, targeting of *Bst1* did not mimic the effects of ageing on ad libitum feeding (Extended Data Fig. 2d, e).

As the aged mouse intestine has also recently been reported to contain reduced Wnt activity[13], we next focused on the extracellular Wnt inhibitor Notum, which was significantly upregulated in old Paneth cells (Fig. 2a, Extended Data Fig. 2c). Notum is a secreted Wnt deacylase that disengages Wnt ligands from Lrp5, Lrp6 and Frizzled receptors and reduces Wnt activity locally during development[15,16]. In the intestine, Wnts are produced by the mesenchymal cells lining the crypt[17,18] and by Paneth cells[2,19]. Wnt ligands produced by the niche adhere to ISC plasma membrane and form a reservoir of stemness-maintaining factors until they become diluted owing to divisions outside the Wnt-producing niche[20]. Of note, the increase in Notum expression was strictly restricted to old Paneth cells (Fig. 2b), in which its secretion could counter the stemness-maintaining function of Wnt ligands. Correspondingly, expression of Wnt-responsive genes was reduced in old Lgr5^{hi} cells (Fig. 2c, Extended Data Fig. 2f). Similarly, in the human intestine, *NOTUM* expression was restricted to Paneth cells, and its expression correlated with age, whereas *LGR5* expression and age showed an inverse correlation (Extended Data Fig. 2g, h).

To test whether Notum indeed affects stemness, we cultured isolated Lgr5^{hi} cells in the absence of Paneth cells and exogenous Wnt ligands. Under these conditions, single Lgr5^{hi} cells form clonal spheroids, whereas more-differentiated cells do not (Fig. 2d). When cells from young mice were treated with biologically active recombinant Notum immediately after isolation to inactivate the membrane-bound Wnts that they were exposed to in vivo, their colony-forming efficacy and the size of formed spheroids were markedly reduced (Fig. 2d, Extended Data Fig. 3a, b). By contrast, colony formation by untreated cells from old mice was already reduced and Notum treatment did not have further effects. Correspondingly, exogenous Wnt ligands increased organoid-forming capacity and long-term regenerative growth specifically in the old crypts (Extended Data Fig. 3c, d). However, exogenously administered Notum had no effect on isolated crypts with tightly connected Paneth and stem cells (Extended Data Fig. 3e–g), which suggests that recombinant Notum could not access the Wnt ligands produced by the Paneth cells. Demonstrating the role of epithelial Wnt secretion, inhibition of Porcupine[20] reduced clonogenic growth and the Lgr5^{hi}:Paneth cell ratio of young organoids similarly to ageing (Extended Data Fig. 3h–j). In sum, these data highlight the consequences of reduced Wnt activity, and suggested that Notum expressed by Paneth cells could represent a mechanism for reducing Wnt activity in the old intestinal epithelium.

Notum is regulated by the canonical Wnt pathway, forming a negative-feedback loop[16]. However, unlike in Lgr5^{hi} ISCs, expression of Wnt-responsive genes was not significantly altered in old Paneth cells (Supplementary Table 1). To find other candidate pathways that regulate Notum in Paneth cells, we performed gene set enrichment analysis (GSEA) and found that transcripts associated with activity of mTORC1 were significantly increased in old Paneth cells (Extended Data Fig. 4a). mTOR signalling is linked with ageing[3], and in the intestine, mTORC1 modulates ISC activity via the Paneth cell niche in response to calorie intake[12]. We detected higher levels of phosphorylation of ribosomal protein S6 (pS6), a downstream effector of mTORC1, in Paneth cells of old mice (Fig. 2e, Extended Data Fig. 4b–d); this increase in pS6 phosphorylation was also reflected in whole-crypt preparations (Extended Data Fig. 4e). However, the frequencies of pS6$^+$ Paneth cells (Extended Data Fig. 4c) or pS6$^+$ crypts (data not shown) were not changed, supporting an increase in mTORC1 activity at the level of single Paneth cells. By contrast, pS6 levels in ISCs were unchanged (Extended Data Fig. 4f, g) but—as reported for the liver[3]—the age-induced mTORC1 activity in Paneth cells was associated with increased body mass (Extended Data Fig. 4h), potentially contributing to increased mTORC1 activity in old Paneth cells.

Inhibition of mTORC1 with rapamycin or by caloric restriction extends lifespan by inducing multisystemic effects[21–23]. When old crypts were transiently treated with rapamycin, regenerative function was restored (Extended Data Fig. 5a, b). Moreover, two-week-long in vivo treatment of old mice with rapamycin resulted in a marked rejuvenation of intestinal regenerative capacity that was attributable to effects on both Paneth cells and ISCs (Extended Data Fig. 5c–h). However, in contrast to caloric restriction[12], systemic rapamycin treatment induced broad changes in expression of intestinal Wnt ligands, including the stromally produced Wnt4, which regulates Notum expression in the developing ovary[24] (Extended Data Fig. 5i). Correspondingly, Notum expression was increased in crypts from

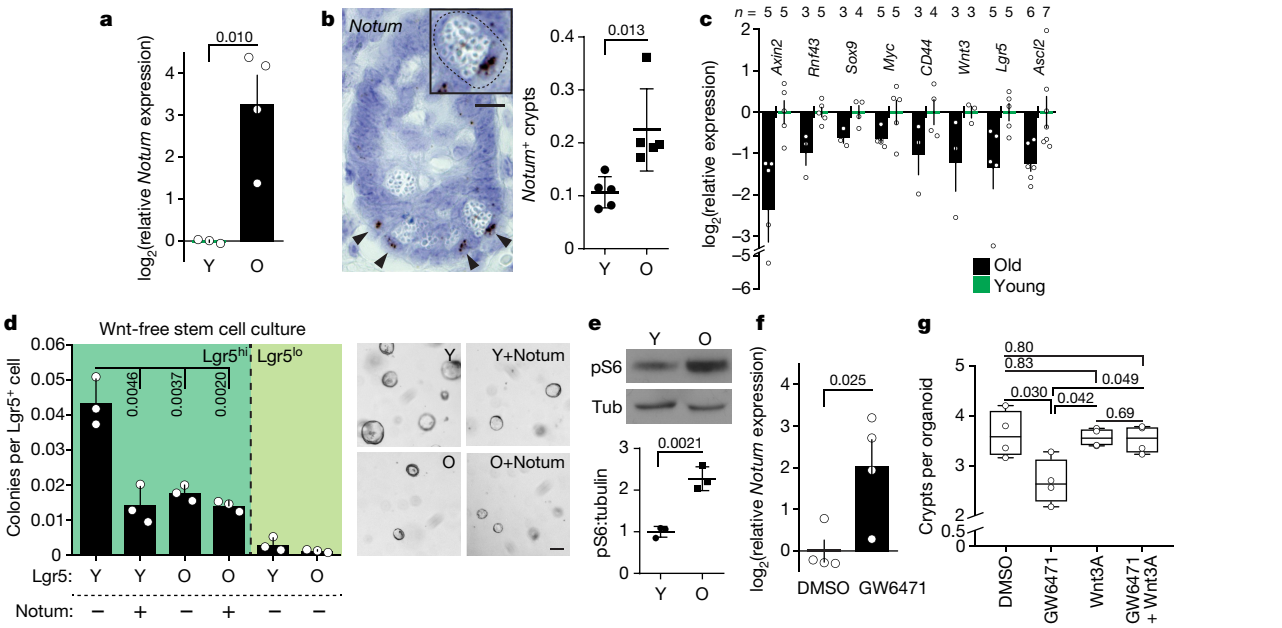


Fig. 2 | Increased *Notum* expression from Paneth cells attenuates Wnt signals in old stem cells. a, Relative *Notum* mRNA expression from isolated Paneth cells ($n = 4$ old, $n = 3$ young mice). **b**, In situ analysis and quantification of *Notum* mRNA expression in mouse jejunum ($n = 5$ mice per age group). Data are mean ± s.d. Scale bar, 10 μm. **c**, Expression of Wnt-responsive genes in isolated old Lgr5^{hi} stem cells relative to young stem cells. Number of mice analysed is shown. **d**, Clonogenic capacity of isolated Lgr5^{+} cells from young and old mice cultured with or without 1 μg ml^{-1} recombinant Notum ($n = 3$ mice per age group). Representative images from day seven of culture. Data are mean ± s.d. Scale bar, 100 μm. **e**, Immunoblot and quantification of lysates from isolated young and old Paneth cells ($n = 3$ mice per age group). Tub, tubulin. Data are mean ± s.d. **f**, Relative *Notum* mRNA expression in small-intestinal organoids of young mice treated for 48 h with 5 μM of the PPAR-α inhibitor GW6471 ($n = 4$ biologically independent samples). **g**, Regenerative growth of small-intestinal organoids at day six. Organoids were treated with DMSO or GW6471 and/or 100 ng ml^{-1} Wnt3A for the first two days ($n = 4$ biologically independent samples). Student's paired *t*-test. Other than in box plots, data are mean ± s.e.m.; two-tailed unpaired Student's *t*-test; *P* values shown in the corresponding panels. $P < 0.05$ is considered significant. For gel source data see Supplementary Fig. 3.

rapamycin-treated mice (Extended Data Fig. 5j), possibly also reflecting an increase in the number of Paneth cells induced by rapamycin[12]. To address the role of mTORC1 activity in the intestinal epithelium in vivo without the rapamycin-induced systemic and stromal effects, we activated mTORC1 specifically in the intestinal epithelium of mice by Villin–Cre-mediated deletion of tuberculosis sclerosis complex 1 (Tsc1). *Tsc1* deletion induced mTORC1 activation and Notum expression in Paneth cells, and reduced organoid-forming capacity (Extended Data Fig. 5k–m). In sum, these results indicate that increased cell-autonomous mTORC1 activity in Paneth cells contributes to the regenerative decline of the old intestinal epithelium.

As mTORC1 does not directly regulate transcription, we searched for factors that mediate Notum expression downstream of mTORC1 activation. To that end, GSEA analysis of old Paneth cells also indicated a significant reduction in expression of genes that are regulated by PPAR-α and PPAR-δ (Extended Data Fig. 6a). mTORC1 activity inhibits PPAR-α[3], and we identified a putative binding site for PPAR-α in the *Notum* gene (Extended Data Fig. 6b). To test whether downregulation of PPAR-α may contribute to the observed ageing phenotypes, we treated young organoid cultures with the PPAR-α antagonist GW6471. GW6471 increased expression of Notum, reduced regenerative growth and decreased the Lgr5^{hi}:Paneth cell ratio (Fig. 2f, g, Extended Data Fig. 6c). Moreover, the ageing-mimicking effects of GW6471 were abrogated by Wnt supplementation (Fig. 2g, Extended Data Fig. 6d). These data indicate that age-associated change in the mTOR–PPAR-α axis modifies Notum expression and the intestinal regenerative capacity in a Wnt-dependent fashion.

Finally, to investigate whether endogenous Notum expression is functionally relevant for the regenerative function, we targeted Notum in organoids to knock out gene function (Extended Data Fig. 7a). Notum-knockout organoids showed increased regenerative capacity in vitro and higher growth rate when orthotopically transplanted to recipient mouse submucosa (Fig. 3a, Extended Data Fig. 7b, c). Moreover, regenerative function of old organoids improved significantly after Notum deletion (Fig. 3b, Extended Data Fig. 7d). Conversely, activation of endogenous Notum expression by CRISPR activation decreased Wnt signalling and colony-forming capacity of CD24^{med}SSC^{lo} cells containing the ISCs of the targeted organoids (Extended Data Fig. 7e–g). Finally, to test whether the regenerative capacity of old intestines can be increased via the intestinal stem cell niche, we used ABC99[25], a small-molecule inhibitor of Notum. ABC99 blunted the effects of exogenous Notum and increased the frequency of Lgr5^{hi} cells in vitro (Extended Data Fig. 8a, b), and in vivo treatment of mice by intraperitoneal injection with 10 mg per kg (body weight) ABC99 had no noticeable adverse effects (Extended Data Fig. 8c). Of note, the Lgr5^{hi} cells that were isolated from old mice after seven days of in vivo treatment with ABC99 demonstrated colony-forming capacity comparable to cells from untreated young mice (Fig. 3c). Moreover, stem-cell-supporting function of Paneth cells was also restored, which suggests autocrine regulation (Extended Data Fig. 8d). To address whether Notum modulates Wnt activity of ISCs in vivo, we next compared the nuclear β-catenin levels of ISCs between Paneth cells to those of more-differentiated transit-amplifying cells that are not in contact with Notum-producing Paneth cells (Extended Data Fig. 8e). As expected, untreated old ISCs had reduced nuclear β-catenin levels (Fig. 3d). ABC99 increased the nuclear β-catenin levels of ISCs specifically in old mice (Fig. 3d, Extended Data Fig. 8f). This increased Wnt activity in old stem cells was reflected in increased proliferation, specifically in Olfm4^{+} stem and progenitor cells, in comparison to more differentiated transit-amplifying cells (Fig. 3e). To formally test whether Notum inhibition promotes regeneration of old intestine, we analysed how advance Notum inhibition affects recovery from 5-fluorouracil (5-FU) chemotherapy-induced mucositis[26,27] that results in loss of body weight owing to compromised water retention and nutrient intake[27]. We treated mice with 100 mg per kg (body weight) 5-FU, as the weight of young mice

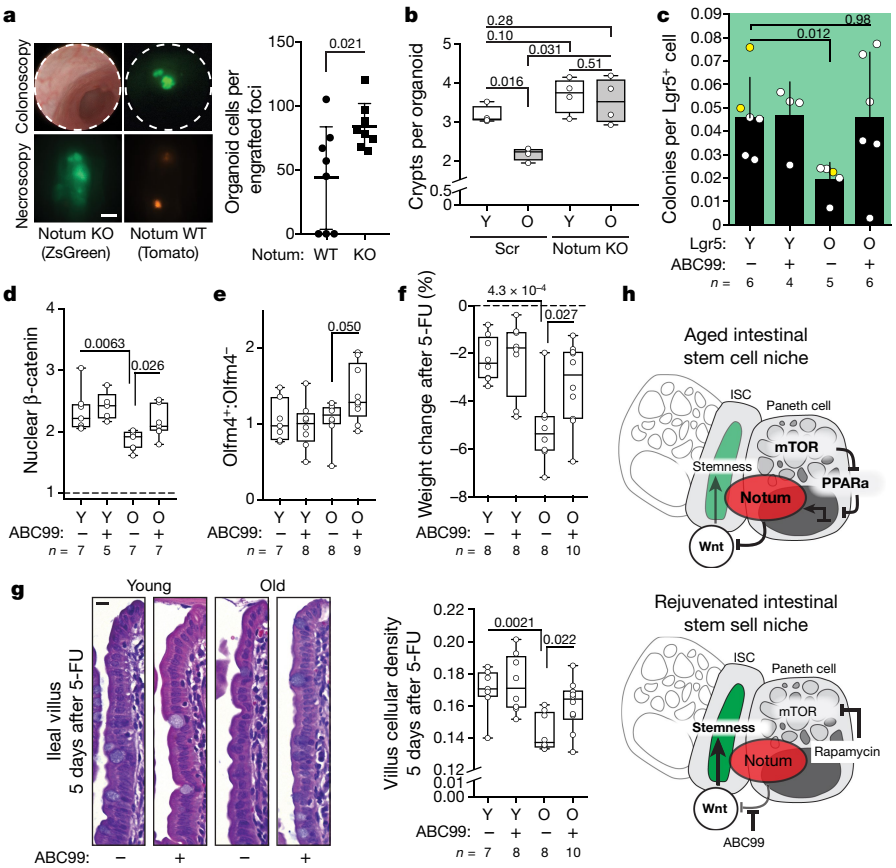


Fig. 3 | Inhibiting Notum activity in vivo restores Wnt-mediated Paneth and stem cell function. a, Growth of CRISPR-targeted organoids after orthotopic transplantation. Notum-knockout (KO) organoids were co-transplanted with competing scramble-targeted controls (Scr). ($n = 8$ mice transplanted). Scale bar, 1 mm. WT, wild type. **b**, Regenerative growth of CRISPR-targeted small-intestinal organoids from young and old mice ($n = 4$ mice per group). Student's paired t-test. **c**, Clonogenic growth of Lgr5^{hi} stem cells from ABC99-treated mice. Control mice received an equal amount of the inactive analogue ABC101 (yellow circles) or vehicle (number of analysed mice shown). **d**, Quantification of relative nuclear β-catenin intensity of crypt base columnar cells (number of analysed mice shown). **e**, Quantification of EdU$^+$ cellular frequencies within the crypt, indicating Olfm4$^+$ cells (number of analysed mice shown). **f**, Average weight change after 5-FU treatment for five days (number of analysed mice shown). **g**, Left, representative images of haematoxylin and eosin (H&E) staining of villi after 5-FU treatment for five days. Scale bar, 10 μm. Right, quantification of cellular density in ileal villi (cells per micrometre; number of analysed mice shown). **h**, Schematic of the model for stem cell maintenance by Paneth cells in the aged niche. Bold weighting of arrows, lines and text indicates a higher strength of the biological signal. In these experiments, old mice were more than 21 months old. Other than in box plots, data are mean ± s.d.; two-tailed unpaired Student's t-test; P values shown in corresponding panels. $P < 0.05$ is considered significant.

recovered fully within 5 days of this dose, but old mice did not recover (Extended Data Fig. 9a). When Notum activity was inhibited with ABC99 for eight days before 5-FU treatment, weight loss in old mice was significantly reduced (Fig. 3f, Extended Data Fig. 9b). Moreover, the density of differentiated cells in the villi was restored to a level similar to that in the young mice, indicating increased regeneration by old stem cells (Fig. 3g). These data demonstrate that Notum produced by Paneth cells attenuates the regenerative capacity of aged intestinal epithelium in vivo by reducing Wnt activity specifically in stem cells.

Appropriate Wnt levels are crucial for many stem cell compartments and alterations are seen in multiple pathologies[28]. Stromal Wnt signals have previously been shown to maintain the epithelial stem cell pool in the absence of Paneth cells under normal tissue homeostasis[17,18,29]. However, recent studies underline the importance of epithelial Wnt signalling in regeneration following injury[30]. Here we find that, during ageing, increased Notum expression in Paneth cells of the ISC niche in mouse and human inhibits Wnt signalling and reduces stem cell maintenance and regeneration. Simultaneously, reversing observed changes in mTORC1–PPAR-α signalling restored epithelial regeneration. Our findings underscore the importance of niche-regulated Wnt signals in promoting stemness and demonstrate a link between ageing-associated metabolic changes and tissue maintenance. Such effects could be missed by studies that demonstrate unaltered clonal dynamics of crypts during ageing[31]. Because Wnt–β-catenin signalling can modulate fatty acid oxidation[32], increasing Wnt activity in old stem cells could also help to restore the age-induced decline in fatty acid oxidation[14]. However, further studies are required to address whether the mechanisms described here also affect tumour risk in the old intestine[33]. In any case, niche–stem cell interactions could provide safer strategies to target tissue renewal and age-related decline than strategies directly targeting stem cells. Activation of PPAR-α or PPAR-δ signalling is not an attractive option in this regard, as PPAR-δ was recently demonstrated to confer tumour-initiating capacity to non-stem cells in the intestine[34]. Notum inhibition with selective inhibitors, such as the ABC99 used here, may represent a safer way to treat gastrointestinal complications and reduce harmful side-effects of chemotherapeutic agents that pose a particular challenge for older individuals[35].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1383-0.

Received: 12 April 2016; Accepted: 10 June 2019;
Published online 10 July 2019.

1. Barker, N. et al. Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* **449**, 1003–1007 (2007).
2. Sato, T. et al. Paneth cells constitute the niche for Lgr5 stem cells in intestinal crypts. *Nature* **469**, 415–418 (2011).
3. Sengupta, S., Peterson, T. R., Laplante, M., Oh, S. & Sabatini, D. M. mTORC1 controls fasting-induced ketogenesis and its modulation by ageing. *Nature* **468**, 1100–1104 (2010).
4. Molofsky, A. V. et al. Increasing p16INK4a expression decreases forebrain progenitors and neurogenesis during ageing. *Nature* **443**, 448–452 (2016).
5. Rossi, D. J. et al. Deficiencies in DNA damage repair limit the function of haematopoietic stem cells with age. *Nature* **447**, 725–729 (2007).
6. Conboy, I. M. & Rando, T. A. Heterochronic parabiosis for the study of the effects of aging on stem cells and their niches. *Cell Cycle* **11**, 2260–2267 (2012).
7. Sato, T. et al. Single Lgr5 stem cells build crypt–villus structures *in vitro* without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
8. Warren, P. M., Pepperman, M. A. & Montgomery, R. D. Age changes in small-intestinal mucosa. *Lancet* **312**, 849–850 (1978).
9. Feibusch, J. M. & Holt, P. R. Impaired absorptive capacity for carbohydrate in the aging human. *Dig. Dis. Sci.* **27**, 1095–1100 (1982).
10. Feldman, M., Cryer, B., McArthur, K. E., Huet, B. A. & Lee, E. Effects of aging and gastritis on gastric acid and pepsin secretion in humans: a prospective study. *Gastroenterology* **110**, 1043–1052 (1996).
11. Potten, C. S., Martin, K. & Kirkwood, T. B. Ageing of murine small intestinal stem cells. *Novartis Found Symp.* **235**, 66–79 (2001).
12. Yilmaz, O. H. et al. mTORC1 in the Paneth cell niche couples intestinal stem-cell function to calorie intake. *Nature* **486**, 490–495 (2012).
13. Nalapareddy, K. et al. Canonical Wnt signaling ameliorates aging of intestinal stem cells. *Cell Rep.* **18**, 2608–2621 (2017).
14. Mihaylova, M. M. et al. Fasting activates fatty acid oxidation to enhance intestinal stem cell function during homeostasis and aging. *Cell Stem Cell* **22**, 769–778 (2018).
15. Giráldez, A. J., Copley, R. R. & Cohen, S. M. HSPG modification by the secreted enzyme Notum shapes the Wingless morphogen gradient. *Dev. Cell* **2**, 667–676 (2002).
16. Kakugawa, S. et al. Notum deacylates Wnt proteins to suppress signalling activity. *Nature* **519**, 187–192 (2015).
17. Shoshkes-Carmel, M. et al. Subepithelial telocytes are an important source of Wnts that supports intestinal crypts. *Nature* **557**, 242–246 (2018).
18. Degirmenci, B., Valenta, T., Dimitrieva, S., Hausmann, G. & Basler, K. GLI1-expressing mesenchymal cells form the essential Wnt-secreting niche for colon stem cells. *Nature* **558**, 449–453 (2018).
19. Farin, H. F., Van Es, J. H. & Clevers, H. Redundant sources of Wnt regulate intestinal stem cells and promote formation of Paneth cells. *Gastroenterology* **143**, 1518–1529 (2012).
20. Farin, H. F. et al. Visualization of a short-range Wnt gradient in the intestinal stem-cell niche. *Nature* **530**, 340–343 (2016).
21. McCay, C. M., Maynard, L. A., Sperling, G. & Barnes, L. L. Retarded growth, life span, ultimate body size and age changes in the albino rat after feeding diets restricted in calories. *Nutr. Rev.* **33**, 241–243 (1975).
22. Harrison, D. E. et al. Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392–395 (2009).
23. Lamming, D. W. et al. Rapamycin-induced insulin resistance is mediated by mTORC2 loss and uncoupled from longevity. *Science* **335**, 1638–1643 (2012).
24. Naillat, F. et al. Identification of the genes regulated by Wnt-4, a critical signal for commitment of the ovary. *Exp. Cell Res.* **332**, 163–178 (2015).
25. Suciu, R. M., Cognetta, A. B., III, Potter, Z. E. & Cravatt, B. F. Selective irreversible inhibitors of the Wnt-deacylating enzyme NOTUM developed by activity-based protein profiling. *ACS Med. Chem. Lett.* **9**, 563–568 (2018).
26. Longley, D. B., Harkin, D. P. & Johnston, P. G. 5-fluorouracil: mechanisms of action and clinical strategies. *Nat. Rev. Cancer* **3**, 330–338 (2003).
27. Song, M. K., Park, M. Y. & Sung, M. K. 5-fluorouracil-induced changes of intestinal integrity biomarkers in BALB/c mice. *J. Cancer Prev.* **18**, 322–329 (2013).
28. Nusse, R. & Clevers, H. Wnt/β-catenin signaling, disease, and emerging therapeutic modalities. *Cell* **169**, 985–999 (2017).
29. Kim, T. H., Escudero, S. & Shivdasani, R. A. Intact function of Lgr5 receptor-expressing intestinal stem cells in the absence of Paneth cells. *Proc. Natl Acad. Sci. USA* **109**, 3932–3937 (2012).
30. Zou, W. Y. et al. Epithelial WNT ligands are essential drivers of intestinal stem cell activation. *Cell Rep.* **22**, 1003–1015 (2018).
31. Kozar, S. et al. Continuous clonal labeling reveals small numbers of functional stem cells in intestinal crypts and adenomas. *Cell Stem Cell* **13**, 626–633 (2013).
32. Frey, J. L. et al. Wnt–Lrp5 signaling regulates fatty acid metabolism in the osteoblast. *Mol. Cell. Biol.* **35**, 1979–1991 (2015).
33. Huels, D. J. et al. Wnt ligands influence tumour initiation by controlling the number of intestinal stem cells. *Nat. Commun.* **9**, 1132 (2018).
34. Beyaz, S. et al. High-fat diet enhances stemness and tumorigenicity of intestinal progenitors. *Nature* **531**, 53–58 (2016).
35. Chang, S., Goldstein, N. E. & Dharmarajan, K. V. Managing an older adult with cancer: considerations for radiation oncologists. *BioMed Res. Int.* **2017**, 1695101 (2017).

METHODS

Isolation of mouse small-intestinal crypts. Mouse small-intestinal crypts were isolated as previously described[12]. In brief, mouse small intestines were flushed with cold PBS and opened, and mucus was removed. The intestine was cut into small fragments and incubated with several changes of 10 mM EDTA in PBS on ice for 2 h. Epithelium was detached by vigorous shaking. To enrich crypts, tissue suspension was filtered through 70-μm nylon mesh. Enriched crypts were washed once with cold PBS and plated into 60% Matrigel (BD Biosciences) with ENR medium. Y-27632 (10 μM) was added to the medium for the first two days.

Isolation of human colonic crypts. Crypts from human colonic biopsies were isolated by vigorous shaking after 1-h incubation in ice cold PBS with 10 mM EDTA. To enrich crypts, tissue suspension was filtered through 70-μm nylon mesh. Enriched crypts were washed once with cold PBS and plated into 60% Matrigel (BD Biosciences) and cultured as previously described[36].

Organoid culture. Crypts were plated (50–200 crypts per 20 μl drop of 60% Matrigel) and overlaid with ENR medium (Advanced DMEM/F12 (Gibco), 1× Glutamax (Gibco), 100 U/ml penicillin and streptomycin, 10 mM Hepes, 1× B-27 (Gibco), 1× N-2 (Gibco), 50 ng/ml of mouse EGF (RnD), 100 ng/ml noggin (Peprotech), 500 ng/ml of R-spondin-1 (RnD), 1 μM *N*-acetyl-L-cysteine (Sigma-Aldrich)). Y-27632 (10 μM) was added for the first two days of culture. Organoid starting frequency was counted after two days of culture unless otherwise stated in the figure legend. Primary organoids were cultured for five-to-nine days, after which regenerative growth (number of de novo crypt domains per organoid) was quantified and organoids were subcultured. Subculturing was performed by mechanically disrupting organoids to single-crypt fragments, which were replated (1:3) with fresh Matrigel. Secondary cultures were confirmed to start from single crypt domains by inspection, and their survival and de novo crypt number were quantified two days after replating. When indicated, ENR medium was supplemented with rapamycin (CST), GW6417 (Tocris), CHIR99021 (BioVision), IWP-2[37] (Sigma) or Wnt3A (RnD). Equal amounts of vehicle (ethanol or DMSO) was used in controls. ENR supplemented with 10 nM gastrin (Sigma-Aldrich), 100 ng/ml Wnt3A (RnD), 10 mM nicotinamide (Sigma-Aldrich), 500 nM A-83-01 (Sigma-Aldrich) and 10 μM SB202190 (Sigma-Aldrich) was used for isolated human colonic crypts[36]. Colonic organoid starting frequency was counted on day seven.

Single-cell sorting and analysis. To isolate single cells, isolated crypts or grown organoids were dissociated in TrypLE Express (Gibco) with 1,000 U/ml of DnaseI (Roche) at 32 °C (90 s for crypts, 5 min at 37 °C for cultured organoids). Cells were washed and stained with the following antibodies: CD31–PE (Biolegend, Mec13.3), CD45–PE (eBioscience, 30-F11), Ter119–PE (Biolegend, Ter119), EpCAM–APC (eBioscience, G8.8) and CD24–Pacific Blue (Biolegend, M1/69), all at 1:500. Finally, cells were resuspended in SMEM medium (Sigma) supplemented with 7-AAD (Life) (2 μg/ml) for live-cell separation. Cells were sorted using a FACSAria II or FACSAria Fusion (BD Biosciences). Sorting strategies: intestinal stem cells, Lgr5–eGFP^{hi}Epcam^{+}CD24$^{med(or\ -)}$CD31^{-}Ter119 $^{-}$CD45^{-}7-AAD^{-}; Paneth cells, CD24^{hi}SideScatter^{hi}Lgr5–eGFP^{-}Epcam^{+}CD31^{-}Ter119^{-}CD45^{-}7-AAD^{-}; enteroendocrine cells, CD24^{hi}SideScatter^{lo}Lgr5–eGFP^{-}Epcam^{+}CD31^{-}Ter119^{-}CD45^{-}7-AAD^{-}. When analysing organoids, eGFP gates were applied directly on the Epcam^{+}CD31^{-}Ter119^{-}CD45^{-}7-AAD^{-} population. Equal numbers of Lgr5^{hi} and Paneth cells were co-cultured with ENR medium supplemented with additional 500 μg/ml of R-spondin-1 (to yield a final concentration of 1 μg/ml), 100 ng/ml Wnt3A and 10 μM of Jagged-1 peptide (Anaspec) for the first six days. Y-27632 (10 μM) was added to the medium for the first four days. Single-cell starting frequency and clonogenic growth of primary organoids were analysed at days 5–9. Long-term organoid-forming capacity was quantified from twice-subcultured organoids 21 days after isolation. Culture of isolated Lgr5^{hi} or CD24^{med}Sidescatter^{lo} cells without Paneth cells and Wnt ligands was performed in ENR supplemented with 10 μM Chir99021, 10 μM Y-27632 and 1 μg/ml recombinant human Notum (RnD) and/or 50–500 nM ABC99[25] when indicated for the first 5 days followed by culture in regular ENR medium. Colony-forming capacity was quantified on day 5 or day 7 as indicated in the figure legends. Cross-sectional area of colonies was quantified from bright-field images taken with an inverted cell culture microscope (Nikon TS100 Eclipse, DS-Qi1Mc camera) on day 7. Paneth cells from mT/mG mouse-derived organoids were isolated as CD24^{hi}SideScatter^{hi}Tomato^{+}Epcam^{+}7-AAD^{-} and co-cultured with freshly isolated Lgr5^{hi} stem cells. Cell population analysis was performed with FlowJo software.

CRISPR–Cas9 gene editing of intestinal organoids. Guide RNAs for the target-gene knockout[38] were designed with the CRISPR design tool (http://crispr.mit.edu). Guides were cloned into lentiCRISPR v2 vector. Lentiviral vectors were produced in 293fT cells (ThermoFisher, R70007) and concentrated with Lenti-X concentrator (Clontech). The 293fT cell line was not authenticated in the laboratory, but tested negative for mycoplasma. Cultured intestinal organoids were exposed to 1 mM nicotinamide for 2 days before they were processed for transduction. Organoids were mechanically disrupted and dissociated to small fragments with TrypLE Express supplemented with 1,000 U/ml DnaseI for 5 min at 32 °C. Fragments were washed once with SMEM medium and resuspended to transduction medium (ENR medium supplemented with 8 μg/ml polybrene (Sigma-Aldrich), 1 mM of nicotinamide, 10 μM Y-27632) and mixed with concentrated virus. Samples were spinoculated 1 h at 600*g* 32 °C followed by 2–4 h incubation at 37 °C, after which they were collected and plated on 60% Matrigel overlaid with transduction medium without polybrene. Two to three days after transduction, infected clones were selected by adding 2 μg/ml of puromycin (Sigma-Aldrich) to the medium. Four days after selection, clones that survived were expanded in normal ENR medium and clonogenic growth was assessed. Knockout was confirmed by three-primer PCR around the gRNA-target site. In experiments comparing young and old gene-edited organoids, organoids were for a maximum of seven days before transduction. LentiCRISPR v2 was a gift from F. Zhang (Addgene plasmid 52961)[39].

Oligonucleotides used for generation of gRNAs: Notum (1), CACCGGGCGGGG CTGCCGTCATTGC, AAACGCAATGACGGCAGCCCCGCCC; Notum (2), CA CCGTCGGCGGTGGTTACTCTTTC, AAACGAAAGAGTAACCACCGCCGAC; Bst-1, CACCGTTCTGGGGGCAAGAGCGCGG, AAACCCGCGCTCTTGCCC CCAGAAC; Scramble (1), CACCGCTAAAACTGCGGATACAATC, AAACGAT TGTATCCGCAGTTTTAGC; Scramble (2), CACCGAAAACTGCGGATACAA TCAG, AAACCTGATTGTATCCGCAGTTTTC. Oligonucleotides used for confirming gene-editing: Notum (1), TATGGCGCAAGTCAAGAGCC, CACGTCGGTGACCTGCAATG, CAAGCCAGGTTGACGGCCT; Notum (2), CGGTTTGGGGATGAGGGTAG, GTCGGCGGTGGTTACTCTTT, GCCAGTCTTTGGAGCTCAT; Bst-1, CCACGGGCTAGAGGAATCAA, GCAAGAGCGCGGTGGAC, CTCAGCAGCGTGGTGTACT.

CRISPR–Cas9 gene activation of intestinal organoids. Lenti-SAM-Cre vector[40] was constructed by assembling five DNA fragments with overlapping ends using Gibson Assembly. In brief, fragments containing sequences corresponding to U6-sgRNA-MS2 (PCR amplified from lenti-sgRNA(MS2)-zeo, Addgene plasmid 61427), the PGK promoter, MS2-p65-HSF1-T2A (PCR amplified from lenti-MS2-P65-HSF1-Hygro, Addgene plasmid 61426), and T2A-Cre were assembled by Gibson assembly into a lentiviral backbone following the manufacturer's guidelines. For short guide RNA (sgRNA) cloning, the Lenti-SAM-Cre vector was digested with BsmBI and ligated with BsmBI-compatible annealed oligonucleotides. sgRNAs were designed using the Cas9 activator tool (http://sam.genome-engineering.org). At least five nucleotides were removed from the 5′ end of candidate sgRNAs to enable use of the SAM system with nuclease-active Cas9[41]. If the first nucleotide in the truncated sgRNA sequence was not a G, an additional nucleotide was removed and replaced with a G to enable efficient expression of the sgRNA from the U6 promoter. Sequence against tdTomato was used as a control[42]. LSL-Cas9-eGFP mouse-derived small-intestinal organoids were infected with Lenti-SAM-Cre-derived virus. Cells with successfully integrated constructs were selected by sorting GFP^{+} cells from organoid cultures. Organoids were grown in ENR containing 3 μM CHIR99021 to avoid selection against Notum expression. Activation of Notum expression was confirmed by quantitative PCR with reverse transcription (RT–qPCR) analysis from whole organoids cultured for two days in ENR without CHIR99021. For assessing the effect of endogenous Notum on stem cells, CD24^{med}Sidescatter^{lo} cells were sorted from organoids cultured for 4–5 days in ENR without CHIR99021.

sgRNA sequences used for generating Lenti-SAM-Cre vectors: Notum (dANotum), GCTGGCCGCGGAGAA; tdTomato (dATom), CGAGTTCGAGATCGA.

RT–qPCR. RNA from crypts, single cells and cultured organoids was isolated by Trizol purification according to the manufacturer's instructions (Life Technologies) using glycogen as co-precipitant (Life Technologies). Full tissue samples were shredded with ceramic beads (Precellys) in RLT buffer and RNA was isolated by RNAeasy+ kit (Qiagen) according to the manufacturer's instructions. Isolated RNA was transcribed with cDNA synthesis kit using OligodT primers (Molecular probes). qPCR amplification was detected by the SYBRGreen (2× SYBRGreen mix, Applied Biosciences) method. Samples were run as triplicates and genes of interest were normalized to *Gapdh* or *Actb*. Primers used for qPCR: *Actb*, CCTCTATGCCAACACAGTGC, CCTGCTTGCTGATCCACATC; *Gapdh*, ATGGTGAAGGTCGGTGTGAA, TGGAAGATGGTGATGGGCTT; *Notum*, CTGCGTGGTACACTCAAGGA, CCGTCCAATAGCTCCGTATG; *Bst1*, ACCCCATTCCTAGGGACAAG, GCCTCCAATCTGTCTTCCAG; *Cd44*, GCACTGTGACTCATGGATCC, TTCTGGAATCTGAGGTCTCC; *Myc*, CAAATCCTGTACCTCGTCCGATTC, CTTCTTGCTCTTCTTCAGAGTCGC; *Ascl2*, CTACTCGTCGGAGGAAAG, ACTAGACAGCATGGGTAAG; *Lgr5*, ACCCGCCAGTCTCCTACATC, GCATCTAGGCGCAGGGATTG; *Axin2*, AGTGCAAACTCTCACCCACC, TCGCTGGATAACTCGCTGTC; *Wnt2b*, CGTGTAGACACGTCCTGGTG, GTAGCGTTGACACAACTGCC; *Wnt3*, TGGAACTGTACCACCATAGATGAC, ACACCAGCCGAGGCGATG; *Wnt4*, GTACCTGGCCAAGCTGTCAT, CTTGTCACTGCAAAGGCCAC; *Wnt5a*,

ATGAAGCAGGCCGTAGGAC, CTTCTCCTTGAGGGCATCG; *Rnf43*, CACCAT AGCAGACCGGATCC, TATAGCCAGGGGTCCACACA; *Sox9*, GAGCCG GATCTGAAGAGGGA, GCTTGACGTGTGGCTTGTTC.

RNA sequencing and data processing. Total RNA from sorted Paneth (four young and five old biological replicates) and Lgr5^{hi} (three young and three old biological replicates) cells were isolated by Trizol purification. Samples were first treated with HL-dsDNase (Articzymes 80200-050) to remove residual DNA. An Ovation Universal RNA-Seq System kit was used for Illumina library preparations (NuGEN Technologies). Purified total RNA (100 ng) was used and primers for ribosomal removal were designed and used as outlined in the kit manual. Libraries were purified with AMPure XP beads (Beckman Coulter), quantified and run on a NextSeq 500 sequencer using 75b single-read kits (Illumina). Adaptor sequences and low-quality reads were removed from the data using cutadapt[43]. The data were mapped to *Mus musculus* genome GRCm38.p4 using STAR[44]. Count data were processed using GenomicFeatures and GenomicAlignments[45], and the differential expression analysis was carried out using DESeq2[46] in R. PreRanked GSEA analysis (http://software.broadinstitute.org/gsea/index.jsp) was performed for fold-change ranked genes with 1,000 permutations[47]. Hallmark, Biocarta and Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets are available via GSEA Molecular Signatures Database. The 'PPARd' gene set was adopted from a previous publication[48]. Gene Ontology enrichment analysis was done for the significantly (adjusted $P < 0.1$) altered genes with Gene Ontology Consortium Enrichment analysis (http://geneontology.org/page/go-enrichment-analysis), using Fisher's exact test not corrected for multiple testing. Subcellular localization for significantly altered genes was taken from the Uniprot database (http://www.uniprot.org/). Putative transcription factor binding sites for PPAR-α on mouse and human *NOTUM* genes were found by using the DECODE database (SABiosciences) and confirmed for mouse using the JASPAR database using PPRE motif (PPARG; RXRA) (http://jaspar.genereg.net). RNA sequencing data from human terminal ileal samples were obtained from the GTEx Portal at the Human Proteome Atlas (https://proteinatlas.org) on 1 June 2018. Sex-matched samples (51 males) were divided into three age groups (20–39, 40–59, >60 years old) and proportions of expression presented. The data are publicly available through ArrayExpress under accession code: E-MTAB-7916 (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7916).

Immunoblotting. Isolated crypts and cells were lysed in RIPA buffer with 1× Halt Protease inhibitor cocktail (ThermoFisher Scientific) and 1× PhosStop (Roche) phosphatase inhibitors. Protein concentrations of cleared lysates were measured by BCA Protein Kit (ThermoFisher Scientific). For sorted cells equal loading was adjusted by sorting the same number of cells. Samples were run on 4–12% Bis-Tris protein gels (Life Technologies) and blotted on nitrocellulose membranes. Membranes were incubated with primary antibodies: pS6 (Ser240/244, CST,5364 for Fig. 2h and Extended Data Figs. 4e, 5d; 1:1,000), pS6 (Ser235/236, CST, 4858; 1:500 for Extended Data Fig. 5k), S6 (CST, 2217; 1:500), H3 (CST, 4499; 1:1,000), β-actin (CST, 4967; 1:2,000), α-tubulin (CST, 2144; 1:1,000) and pS6K (ImmunoWay,YP0886; 1:500) at 4 °C overnight, and HRP-conjugated anti-rabbit (Sigma-Aldrich; 1:5,000) or anti-mouse (CST; 1:1,000) for 1 h at room temperature. Signal was detected using ECL reagent Supersignal West Femto (ThermoFisher Scientific). Densitometry was performed with ImageJ, normalizing to β-actin or α-tubulin.

Immunohistochemistry and immunofluorescence. Tissues were fixed in 4% PFA, paraffin-embedded and sectioned. Antigen retrieval was performed by boiling in pH 6 citrate buffer (Sigma-Aldrich) for 5 min. Antibodies: lysozyme (DAKO, EC3.2.1.17; 1:750), Ki67 (Abcam, ab15580; 1:300), pS6 (Ser240/244) (CST, 5364; 1:1,000), Olfm4 (clone PP7, a gift from CST (Extended Data Fig. 1k), CST, 39141 (Extended Data Fig. 8g); 1:300), β-catenin (BD, 610153; 1:300), E-cadherin (BD, 610181; 1:500). Antigen retrieval was followed by permeabilization with 0.5% Triton-X100 (Sigma) and, in the case of analysis of EdU incorporation, was followed by EdU Click-IT chemistry according to manufacturer's instructions (ThermoFisher Scientific). Primary antibodies were detected with biotin-conjugated secondary antibodies and DAB substrate on peroxidase-based system (Vectastain Elite ABC, Vector Labs). For immunofluorescence, Alexa Fluor 488-, Alexa Fluor 594-, Alexa Fluor 633- and Alexa Fluor 647-conjugated anti-rabbit or anti-mouse secondary antibodies (Life Technologies, all 1:500) were used. Nuclei were co-stained with DAPI (Life Technologies, 1 μg/ml) or Hoechst 33342 (Life Technologies, 1 μg/ml).

Immunocytochemistry. Sorted cell populations were treated as previously described[12]. In brief, they were either centrifuged onto charged microscope slides with a Shandon Cytospin 4 (ThermoFisher) for 3 min at 800 r.p.m. or allowed to settle on poly-L-lysine-coated coverglass-bottomed MatTek dishes for 15 min at 37 °C followed by fixation with 4% PFA and immunostaining. Antibodies: lysozyme (DAKO, EC3.2.1.17; 1:500), Muc2 (SantaCruz,H-300; 1:50), counterstains: Hoechst 33342 (Life Technologies, 1 μg/ml), Phalloidin-647 (Life Technologies; 1:50).

Quantification of nuclear β-catenin localization. Three-micrometre-thick confocal sections of β-catenin-stained ileal segments were captured with a Leica SP5IIHCS confocal microscope and 63× water-immersion objective and 12-bit image colour depth. β-catenin mean fluorescence intensity was measured by a blinded investigator from three nuclear regions of interest of cells in the transit-amplifying zone (cell position +6 and above relative to crypt bottom) followed by measurement of intensities in the nuclei of crypt base columnar cells (CBCs) and Paneth cells (identified by nuclear morphology and cellular shape). Nuclear intensities of CBCs and Paneth cells were always normalized to transit-amplifying cells from the same image.

RNA in situ hybridization. RNA in situ hybridization was performed with RNAScope 2.5HD Assay–Brown according to the manufacturer's protocol (RNAScope ACDBio). Probes used: mouse *Notum*: Mm-Notum 428981; mouse (positive control) *Lgr5*: Mm-Lgr5 312178; Human *NOTUM*: Hs-NOTUM 430311; human *UBC* (positive control): Hs-UBC 310041; *Dapb* (negative control): Probe-DapB 310043.

Of note, human *NOTUM* was only detected from samples freshly fixed in 4% PFA followed by paraffin-embedding. Pathological samples fixed in 10% NBF were not compatible with this probe.

Statistical analysis. No statistical method was used to calculate the sample size. For analysis of in vitro organoid cultures, investigators were blinded when possible, but owing to features of co-culture experiments this was not always possible. Blinded investigators performed all histological quantification. Microsoft Excel v.16.16.8 and Graphpad Prism v.8.0.0 were used for statistical analysis and visualization of data. All data were analysed by two-tailed Student's *t*-test, except RNA sequencing data (see 'RNA sequencing and data processing'), exact *P* values are presented in the corresponding figures. A paired *t*-test was applied if the day of organoid growth quantification varied between pairs (samples processed the same day were paired) or phenotype after treatment was compared to the control from the same mouse (samples from the same mouse were paired). Whether a test was paired or unpaired is noted in the figure legends. $P < 0.05$ was considered significant.

Human biopsy samples. Human ileal and colon tissue biopsies were obtained from 24 healthy subjects who were undergoing a routine colonoscopy. Human jejunal samples were obtained from patients undergoing Roux-en-Y gastric bypass surgery and fixed in 4% PFA before a routine paraffin-embedding protocol. The specimens used for organoid functional assay were stored in normal saline on ice until analysis. Exclusion criteria included any history of malignancy, chronic liver disease, history suggesting a malabsorption disorder, previous intestinal surgery, renal disease, bleeding disorder that would preclude biopsy, active infection or systemic inflammatory disorder. The study regarding relevant samples and associated ethical regulations were approved by the institutional review board of Massachusetts General Hospital and Helsinki University Hospital. Written and informed consent was obtained before enrolment.

Mice. *Lgr5-eGFP-IRES-creERT2* mice[1] were maintained in a C57BL/6J background. *Rosa26^{mTmG}* (JAX 007576), *Tsc1$^{fl/fl}$* (ref. [49]) (*Tsc1$^{tm1Djk/J}$*, JAX 005680), *Rosa26$^{LSL\text{-}ZsGreen}$* (JAX 007906), *Rosa26$^{LSL\text{-}TdTomato}$* (JAX 007909), *Rag2$^{-/-}$* (*B6(Cg)-Rag2tm1.1Cgn/J*, JAX 008449) and *Rosa26$^{LSL\text{-}Cas9\text{-}eGFP}$* (JAX 024857) mice were obtained from Jackson Laboratories and were on mixed background. *Villin-creERT2* mice were a gift from S. Robine and have previously been described[50]. All animal housing and experiments were done under local institutional regulations. Mice were allocated to experimental groups randomly, but without proper randomization. Investigators were not blinded owing to the apparent phenotype of aged mice. For in vivo proliferation analysis, 10 mg/kg of EdU (Sigma) in PBS was injected intraperitoneally 2 h before mice were euthanized. For in vivo Tsc1 deletion, *Villin-creERT2;Tsc1$^{fl/fl}$* mice were given 5 intraperitoneal injections of 100 mg/kg tamoxifen (Sigma) on alternate days. Rapamycin treatment was performed as previously described[12]. ABC99 was produced as previously described[25]: 33.3 mg/ml stock solution in ethanol was prepared freshly and further mixed 1:1:1:17 into Tween-80 (Sigma), PEG-400 (HamiltonResearch) and 0.9% NaCl. Mice were injected intraperitoneally with 10 mg/kg ABC99 daily with a last dose together with 10 mg/kg EdU 2 h before they were euthanized. Control mice were treated with vehicle or an equal amount of the inactive control compound ABC101[25]. 5-FU (Sigma) was reconstituted in DMSO at 100 mg/ml and a single intraperitoneal injection was given to mice with a dose of 100–200 mg/kg (as described in the figure legends). Mice over 24 months of age were considered old, and mice between 3 and 9 months of age were considered young (denoted 'O' and 'Y', respectively, throughout the figure legends), with the exception of Fig. 3f, g, Extended Data Fig. 9a, b and Supplementary Fig. 2, in which old mice were 20–22 months of age. Both sexes were used in all experiments. All animal experiments were approved and carried out in accordance with the guidelines of the Finnish National Animal Experimentation Board and the Committee on Animal Care at MIT.

Organoid transplantation. Notum wild-type and Notum knockout intestinal organoids were generated using *Notum* (2) guide RNAs, as described above, in *Villin-creERT2;Rosa26$^{LSL\text{-}ZsGreen}$* and *Villin-creERT2;Rosa26$^{LSL\text{-}tdTomato}$*

intestinal organoids cultured with 4-OHT to generate zsGreen$^+$ wild-type, zsGreen$^+$ knockout, tdTomato$^+$ wild-type and tdTomato$^+$ knockout organoids. Organoids were grown in Matrigel and cultured with crypt medium. Before transplantation, ZsGreen$^+$ knockout and tdTomato$^+$ wild-type (and, in parallel, ZsGreen$^+$ wild-type and tdTomato$^+$ knockout) organoids were chemically dissociated using Cell Recovery Solution (Corning, 354253), and then resuspended in a 1:1 ratio in 90% crypt medium and 10% Matrigel at a concentration of 25 organoids per μl. Organoids were orthotopically transplanted into the colonic submucosa of *Rag2*$^{-/-}$ recipient mice, as previously described[51,52]. The average volume of each injection was 60 μl. Eight weeks later, engrafted organoids were assessed using fluorescence colonoscopy followed by fluorescence microscopy using GFP and tdTomato filters. Tissues were then fixed in 4% paraformaldehyde for 4–6 h, cryopreserved with 30% sucrose in PBS overnight, and then frozen in OCT. Frozen tissue sections were stained with DAPI to visualize nuclei, and then imaged for tdTomato and GFP. The total number of tdTomato$^+$ and GFP$^+$ cells per mouse was then counted using Fiji.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

RNA sequencing data are publicly available through ArrayExpress with accession code E-MTAB-7916. Source Data for Fig. 1–3 and Extended Data Figs. 1–9 are available with the online version of the paper. All other data are available from the corresponding author upon reasonable request.

36. Sato, T. et al. Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* **141**, 1762–1772 (2011).
37. Chen, B. et al. Small molecule-mediated disruption of Wnt-dependent signaling in tissue regeneration and cancer. *Nat. Chem. Biol.* **5**, 100–107 (2009).
38. Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
39. Sanjana, N. E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
40. Konermann, S. et al. Genome-scale transcriptional activation by an engineered CRISPR–Cas9 complex. *Nature* **517**, 583–588 (2015).
41. Dahlman, J. E. et al. Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat. Biotechnol.* **33**, 1159–1161 (2015).
42. Sánchez-Rivera, F. J. et al. Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* **516**, 428–431 (2014).
43. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 17.1.2001 (2011).
44. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
45. Lawrence, M. et al. Software for computing and annotating genomic ranges. *PLOS Comput. Biol.* **9**, e1003118 (2013).
46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
47. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102**, 15545–15550 (2005).
48. Adhikary, T. et al. Genomewide analyses define different modes of transcriptional regulation by peroxisome proliferator-activated receptor-β/δ (PPARβ/δ). *PLoS ONE* **6**, e16344 (2011).
49. Kwiatkowski, D. J. et al. A mouse model of TSC1 reveals sex-dependent lethality from liver hemangiomas, and up-regulation of p70S6 kinase activity in Tsc1 null cells. *Hum. Mol. Genet.* **11**, 525–534 (2002).
50. el Marjou, F. et al. Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis* **39**, 186–193 (2004).
51. Roper, J. et al. Colonoscopy-based colorectal cancer modeling in mice with CRISPR-Cas9 genome editing and organoid transplantation. *Nat. Protoc.* **13**, 217–234 (2018).
52. Roper, J. et al. *In vivo* genome editing and organoid transplantation models of colorectal cancer and metastasis. *Nat. Biotechnol.* **35**, 569–576 (2017).

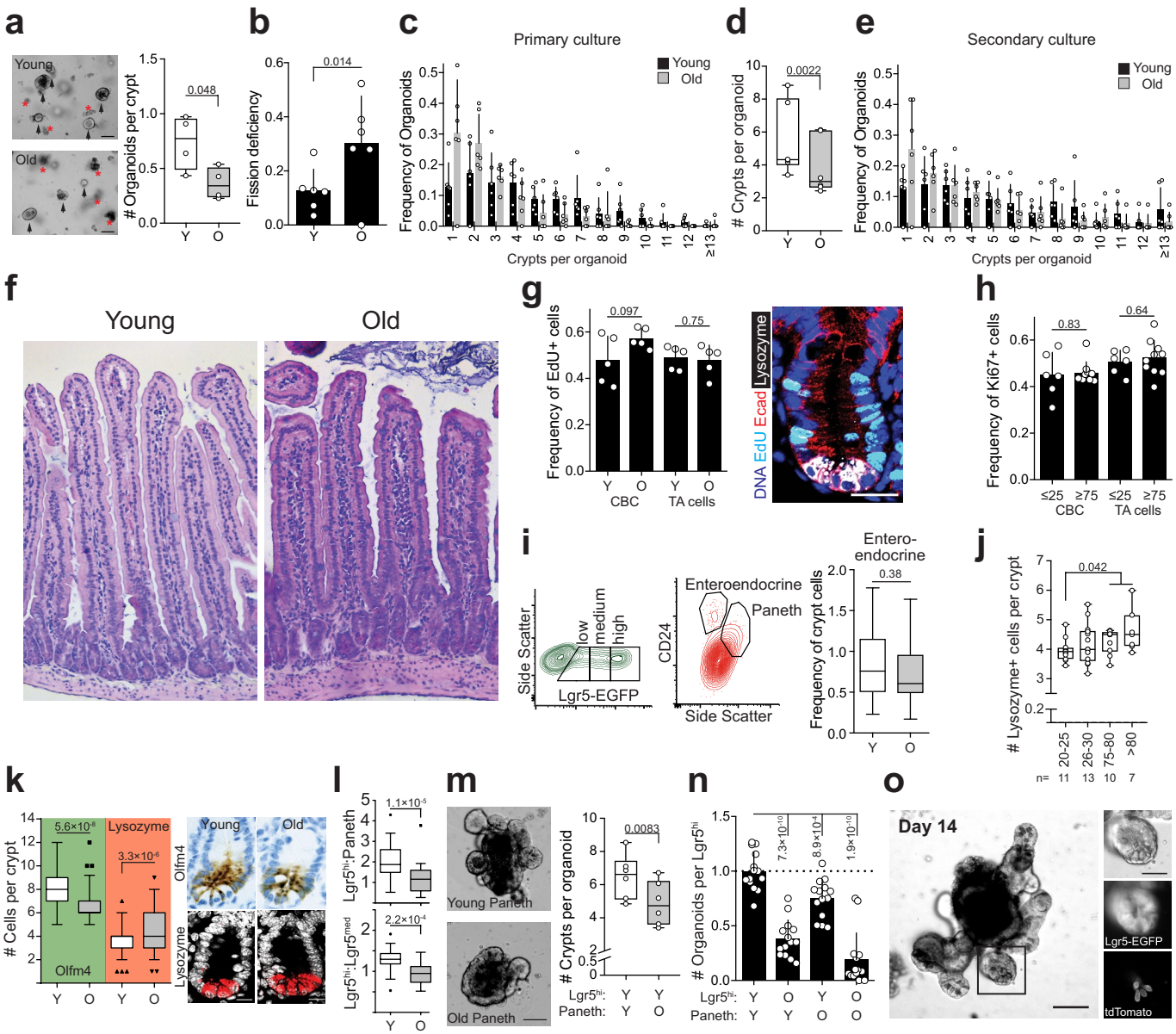
Acknowledgements This study was supported by the Academy of Finland (Research Fellow and Centre of Excellence, MetaStem), Marie Curie CIG (618774), ERC-STG (677809), Swedish Research Council 2018-03078, Sigrid Juselius Foundation, Center for Innovative Medicine, and Wallenberg Academy Fellows program to P.K. N.P. was supported by the Integrative Life Science Doctoral program and by the Research foundation of University of Helsinki. B.F.C. was supported by grant R35 CA231991. We thank the personnel of the DNA sequencing and genomics laboratory for performing the RNA sequencing assays. We thank J. Bärlund, A. Sola-Carvajal, M. Simula and A. Kegel for technical assistance.

Author contributions N.P. and P.K. designed and interpreted the results of all experiments. N.P., S.I., M.M., S.A., A.B.C. III, R.M.S., K.L., E.M., S.G., S.N. and T.T. performed all experiments and analysed the results. J.R. performed and analysed organoid transplantations. O.-P.S. processed and analysed the RNA sequencing data with the help of P.A., P.K. and N.P. N.P., M.M., K.L., E.M., S.A. and S.G. performed and analysed the immunohistochemistry, immunofluorescence and RNA in situ hybridizations. N.G., T.S., A.J., K.P. and A.R. provided the human biopsy material. Ö.H.Y., D.M.S., T.T., T.J. and B.F.C. participated in the design and interpretation of experiments. P.K. and N.P. wrote the paper.

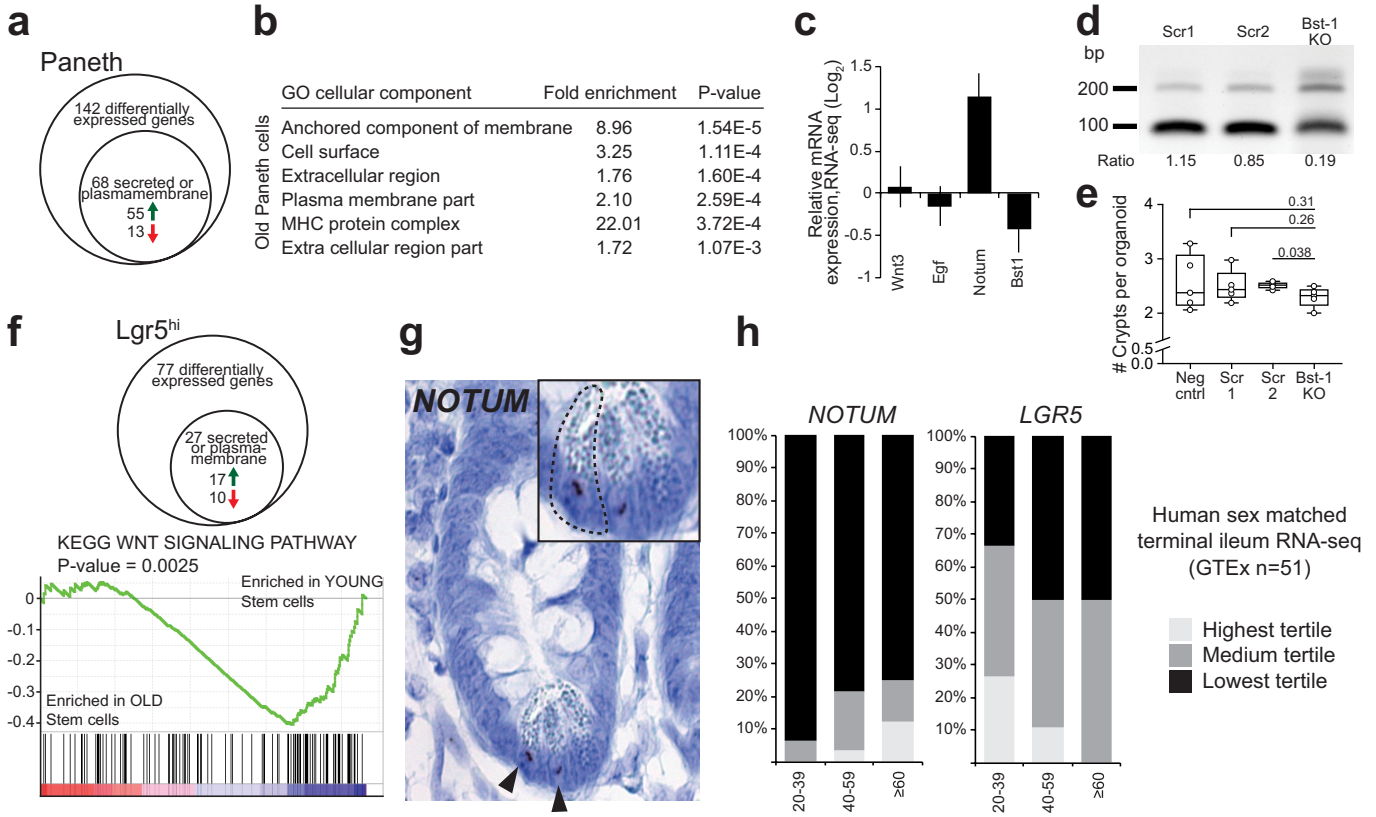
Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1383-0.

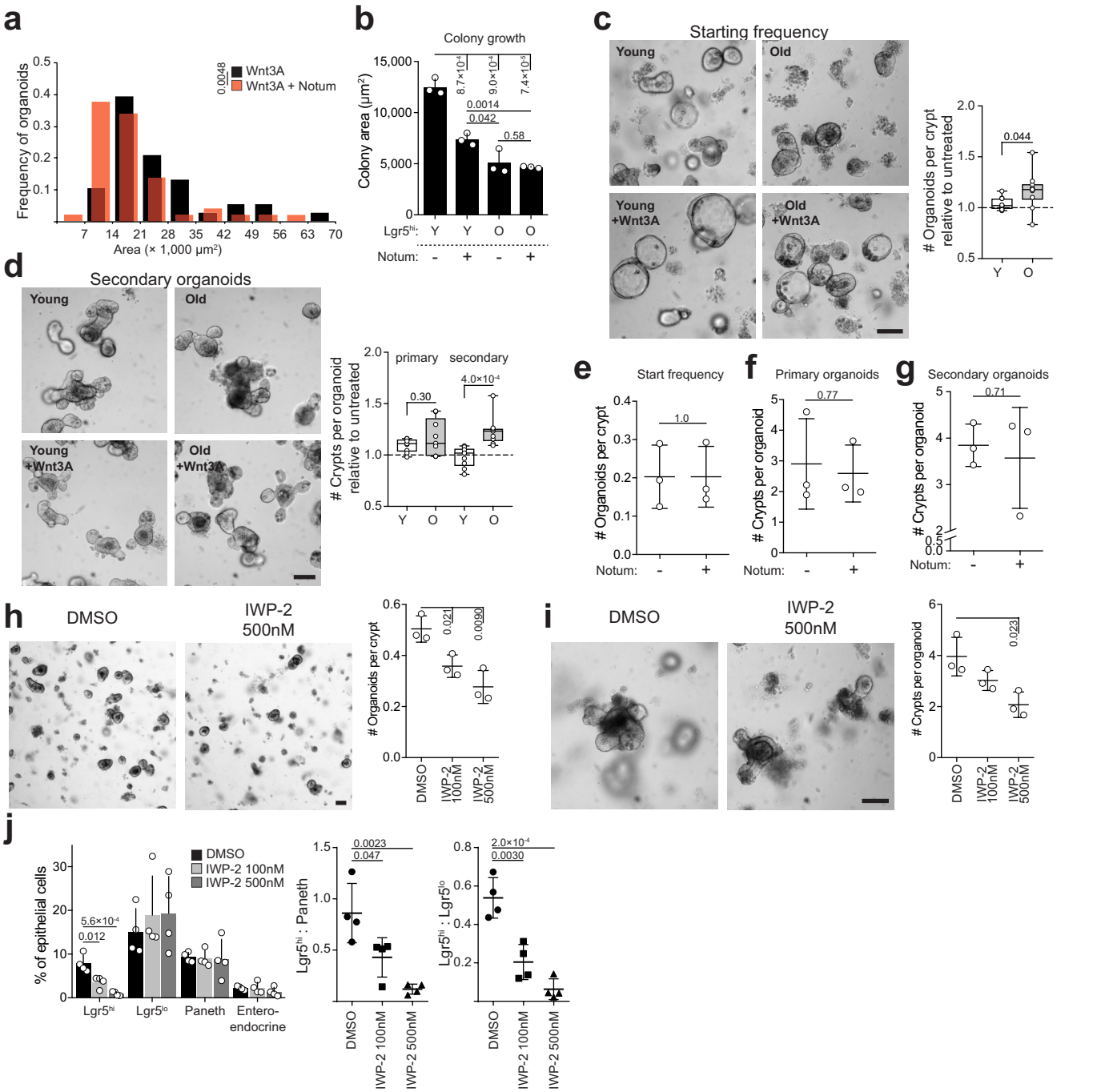
Correspondence and requests for materials should be addressed to P.K.



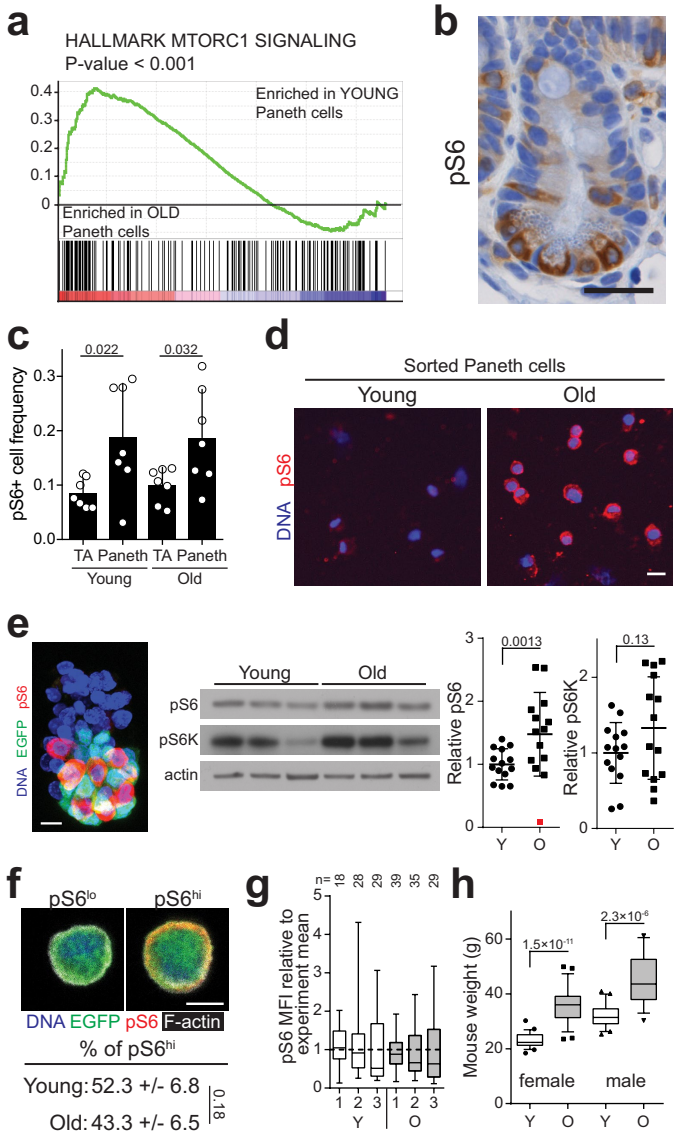
Extended Data Fig. 1 | Characterization of aged intestine. a, Organoid-forming capacity of crypts from young and old mice ($n = 4$ mice per group). Student's paired t-test. **b**, Frequency of organoids unable to form new crypts (fission deficiency) in young and old mice ($n = 6$ mice per group) analysed 5–9 days after isolation. Student's paired t-test. **c**, Distribution of regenerative growth capacity of primary organoids from young and old mice ($n = 6$ mice per group). **d**, Regenerative growth of subcultured secondary mouse organoids ($n = 6$ mice per group). Student's paired t-test. **e**, Distribution of regenerative growth capacity of subcultured secondary organoids from old and young mice ($n = 6$ mice per group). **f**, Representative H&E staining of mouse jejunal sections from young and old mice (four mice analysed per group). **g**, Quantification of EdU$^+$ cells in jejunal crypts 2 h after administration. Only cells next to lysozyme$^+$ Paneth cells were quantified as CBC stem cells. Crypt cells that were not touching lysozyme$^+$ cells were quantified as transit-amplifying (TA) cells ($n = 5$ mice per group). Representative image of crypt stained for EdU (cyan), DAPI (nuclei, blue), lysozyme (white) and E-cadherin (red). Scale bar, 20 μm. **h**, Quantification of Ki67$^+$ cells in human ileal biopsies. Cells at the crypt bottom with elongated nuclei next to postmitotic Paneth cells were counted as CBCs. Cells not at the crypt base were considered transit-amplifying cells. ($n = 6$ for 20–25-year-old donors, $n = 10$ for ≥ 75-year-old donors). **i**, Representative gating of Lgr5^{hi}, Lgr5^{med}, Lgr5^{lo}, Paneth and enteroendocrine cells (in relation to Fig. 1c). Quantification of enteroendocrine cells ($n = 30$ young, $n = 26$ old). For FACS gating strategy, see Supplementary Fig. 1. **j**, Analysis of human ileal biopsy material for lysozyme$^+$ Paneth cells (n values for analysed samples shown). **k**, Immunostaining and quantification of Olfm4$^+$ stem and progenitor cells (green background, $n = 75$ crypts from young and old. Five individuals per age group) and lysozyme$^+$ Paneth cells in jejunal crypts (red background, $n = 115$ crypts from young and $n = 117$ crypts from old mice, five individuals per age group). Whiskers plotted according to Tukey's method. Scale bars, 10 μm. **l**, Ratio of Lgr5^{hi} stem cells and Lgr5^{med} progenitor cells and ratio of Lgr5^{hi} stem cells and Paneth cells analysed by flow cytometry from isolated crypts ($n = 30$ young, $n = 26$ old). Whiskers plotted according to Tukey's method. **m**, Regenerative growth of young Lgr5^{hi} stem cells co-cultured with young or old Paneth cells. Quantification at day 8–11 ($n = 6$). Representative images are from day 8. Scale bar, 100 μm. Student's paired t-test. **n**, Long-term clonogenicity of young and old Lgr5^{hi} stem cells co-cultured with young and old Paneth cells. Serially passaged organoids were quantified 21 days after initial plating ($n = 14$ mice per age group). Combinations compared to average of young Lgr5^{hi} cells co-cultured with young and old Paneth cells. **o**, Fourteen-day co-culture of Paneth cells from tdTomato-expressing mouse (R26-mTmG) with Lgr5^{hi} stem cells from *Lgr5-eGFP-IRES-creERT2* mouse show long-term niche interactions in organoid culture. Scale bar, 100 μm. Similar results were seen in three replicate wells from co-cultures of the same mice. Y, mice between 3 and 9 months of age; O, mice over 24 months of age in all experiments. In box plots, unless otherwise indicated, the line represents median, the box shows interquartile range and whiskers represent the range. All other data are mean ± s.d.; two-tailed unpaired Student's t-test; exact P values shown in corresponding panels.



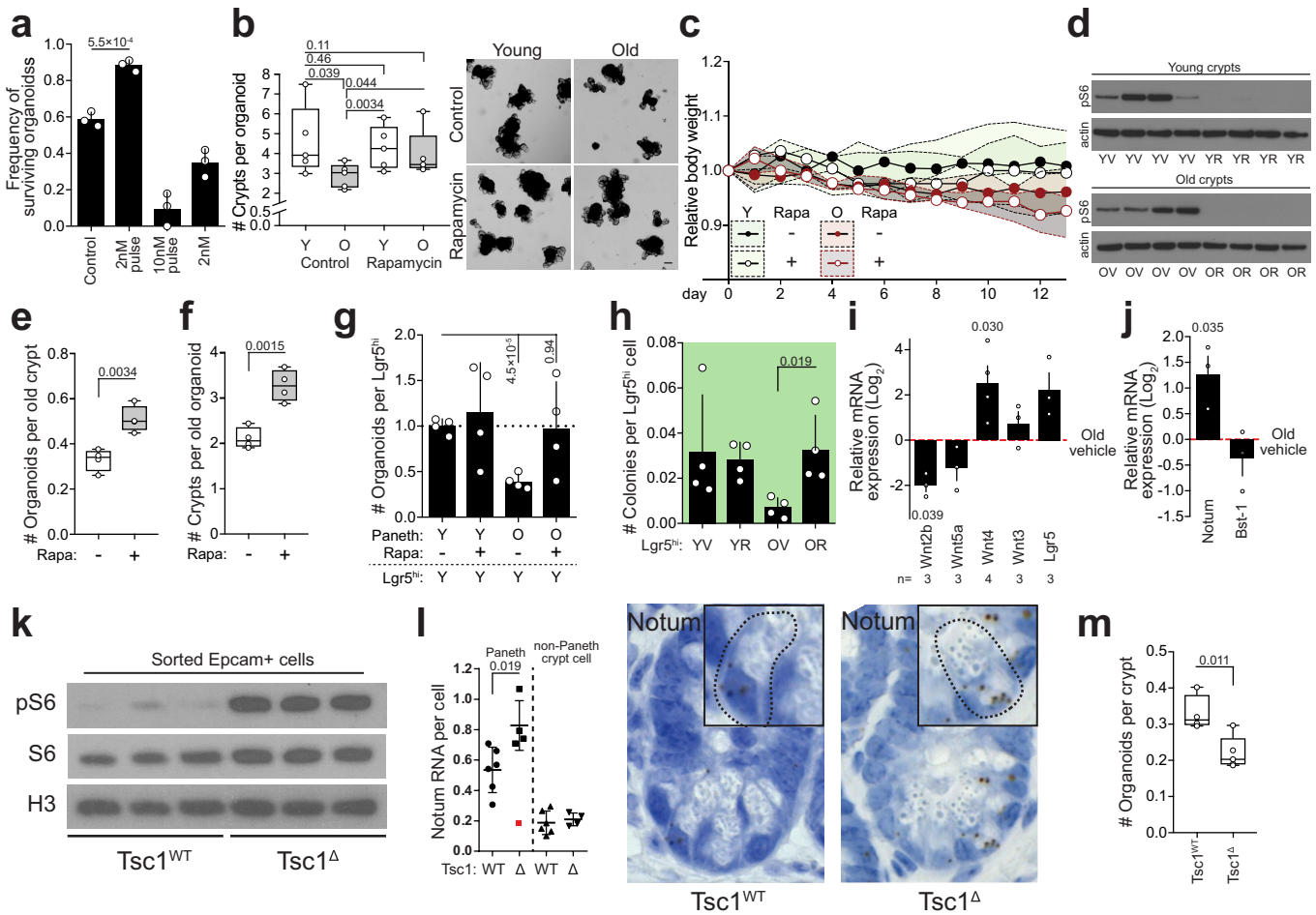
Extended Data Fig. 2 | Characterization of gene expression in old Paneth and ISCs. a, Venn diagram of gene-expression changes in old Paneth cells. ($n = 5$ mice in old, $n = 4$ mice in young) **b**, List of Gene Ontology (GO) terms with the highest enrichment among genes deregulated in old Paneth cells. Fisher's exact test, no correction for multiple testing. **c**, Expression of stem-cell-maintaining factors *Wnt3* and *Egf*, and of *Notum* and *Bst1* in old Paneth cells (RNA sequencing (RNA-seq)). Values show fold change in comparison to young Paneth cells. ($n = 5$ mice in old, $n = 4$ mice in young). **d**, Gene editing of *Bst1* confirmed by PCR strategy with primers flanking the editing site (191-bp product) and hitting the edited site (89-bp product). Representative agarose gel image is shown. Experiment repeated once to validate the organoid line used in **e**. **e**, Regenerative growth of Bst-1 knockout intestinal organoids. Organoids were quantified two days after subculturing ($n = 5$ repeated experiments with the same organoid line). **f**, Venn diagram of gene-expression changes in old Lgr5^{hi} stem cells. GSEA preranked analysis of old versus young Lgr5^{hi} stem cells for the gene list 'KEGG WNT signalling pathway'. Nominal *P* value is shown ($n = 3$ mice per age group). **g**, RNA-scope for *NOTUM* mRNA (brown) in human jejunal section. Expression seen exclusively in Paneth cells (arrowheads and inset). Experiment repeated twice with similar results in independent samples. **h**, Expression of human *NOTUM* and *LGR5* from terminal ileal samples of GTEx Consortium ($n = 51$ sex-matched samples). Expression range is divided into three equal-sized tertiles. In box plots, unless otherwise indicated, the line represents median, the box shows interquartile range and whiskers represent the range. All other data are mean ± s.d.; two-tailed unpaired Student's *t*-test; exact *P* values shown in corresponding panels. For gel source data, see Supplementary Fig. 3.



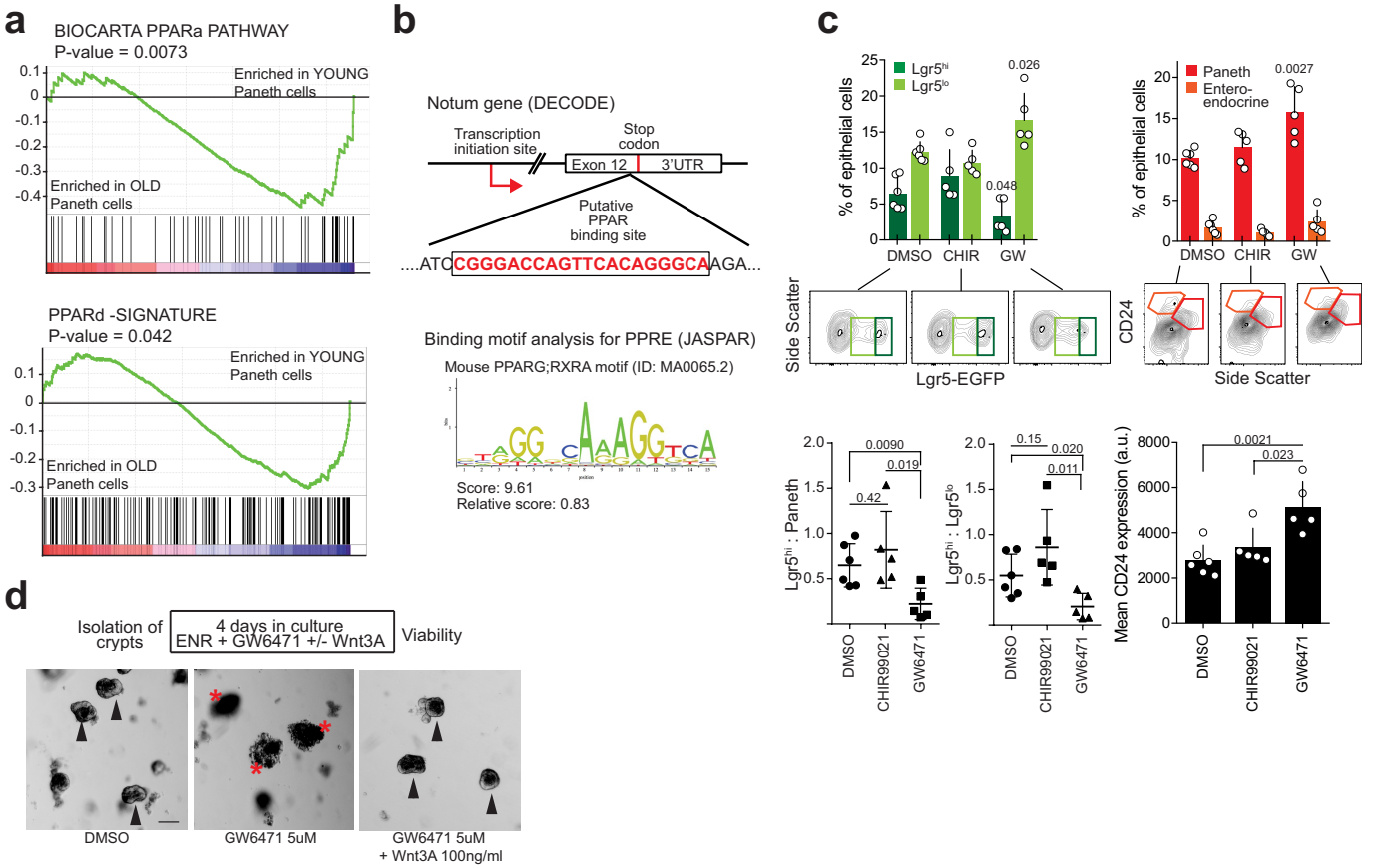
Extended Data Fig. 3 | Wnt ligands increase regenerative capacity of ISCs. a, Distribution of organoid size on day 5 in ENR + 100 ng ml^{-1} Wnt3A ± 1µg ml^{-1} recombinant Notum ($n = 50$ organoids for Notum-treated (red), $n = 38$ organoids for untreated (black)). **b**, Area of colonies from sorted Lgr5^{hi} stem cells from young and old mice ($n = 3$ mice per age group). Area quantified at day 7. **c**, Organoid-forming capacity of crypts from young and old mice treated with 100 ng ml^{-1} Wnt3A. Starting frequency was quantified on day 2 and is represented relative to untreated control ($n = 10$ mice per age group). **d**, Primary and secondary regenerative growth of young and old organoids treated with or without 100 ng ml^{-1} Wnt3A for the first 2 days of culture. Primary organoids were quantified at day 6 and secondary organoids two days after subculturing. Data are represented relative to untreated control ($n = 9$ mice per age group). **e**, Organoid-forming capacity of isolated crypts from young mice treated with or without 1 µg ml^{-1} recombinant Notum ($n = 3$ mice). **f**, Primary regenerative growth of organoids from young mice treated with or without 1 µg ml^{-1} recombinant Notum, quantified on day 6 ($n = 3$ mice). **g**, Secondary regenerative growth of organoids from young mice treated with or without 1 µg ml^{-1} recombinant Notum, quantified on day 2 after subculture ($n = 3$ mice). **h**, Organoid-forming capacity of isolated crypts at day 2 from young mice treated with Porcupine inhibitor IWP-2[37] ($n = 3$ mice). **i**, Primary regenerative growth of organoids treated with IWP-2 for the first two days of culture. Organoids were quantified on day 6 ($n = 3$ mice). **j**, Flow cytometry analysis of cellular frequencies from Lgr5–eGFP organoids two days after treatment with IWP-2 ($n = 4$ mice). Y, mice between 3 and 9 months of age; O, mice over 24 months of age in all experiments. In box plots, unless otherwise indicated, the line represents median, the box shows interquartile range and whiskers represent the range. All other data are mean ± s.d.; two-tailed unpaired Student's *t*-test; exact *P* values shown in corresponding panels.



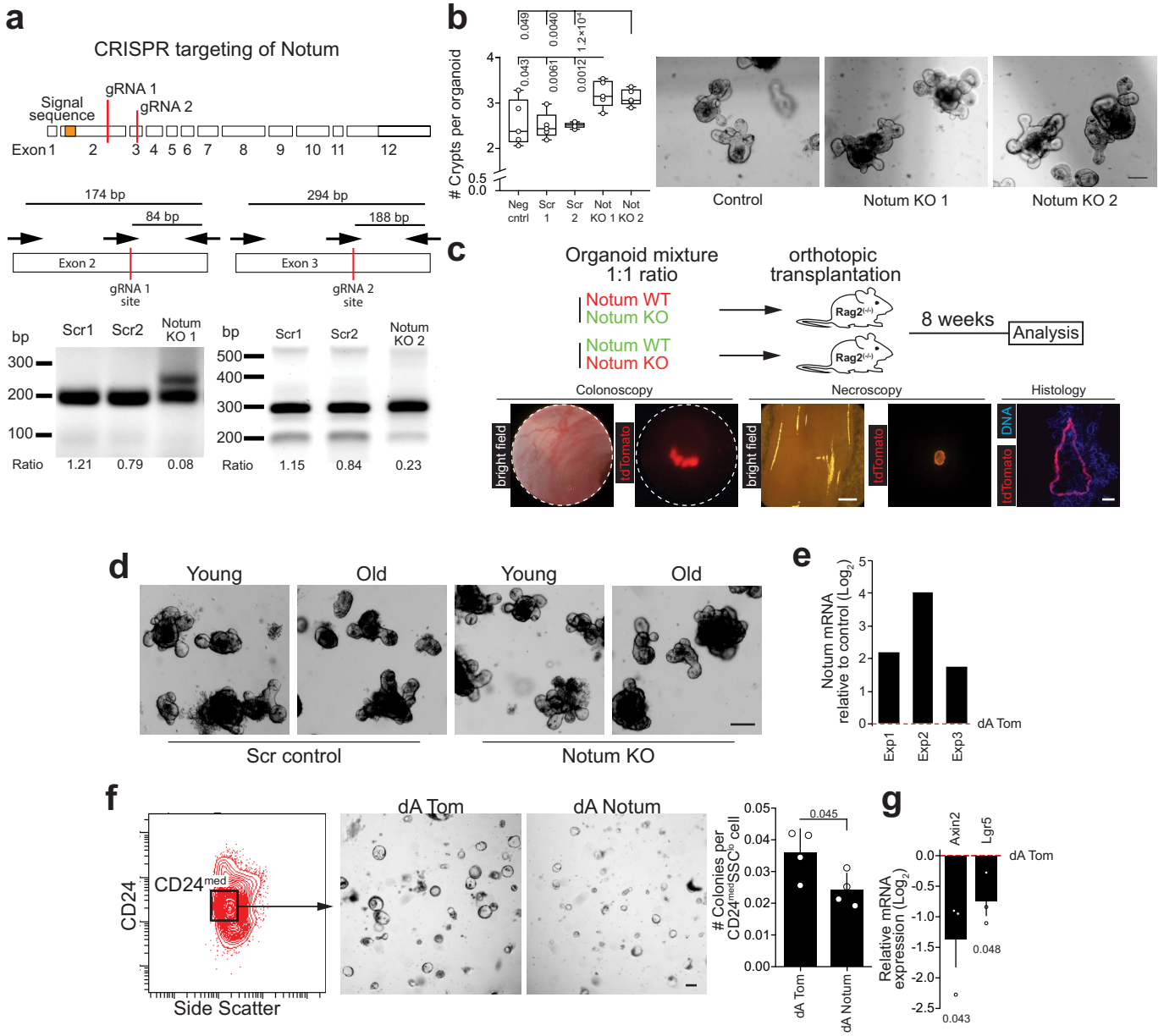
Extended Data Fig. 4 | Increased mTORC1 activity in old Paneth cells but not in ISCs. a, GSEA analysis for gene list 'Hallmark mTORC1' (for statistics, see 'RNA sequencing and data processing' in Methods). Nominal *P* value is shown ($n = 5$ mice in old, $n = 4$ mice in young). **b**, Immunohistochemical staining of pS6 (Ser240/244) at mouse jejunal crypt. pS6^{+} Paneth cells at the crypt bottom are separated by pS6^{-} CBCs. Scale bar, 25 μm. Experiment was repeated for 14 mice with similar results. **c**, Quantification of pS6^{+} cells in jejunal crypts ($n = 7$ mice per age group). **d**, Isolated Paneth cells from young and old mice, stained with pS6 antibody (red), DAPI (nuclei, blue). Scale bar, 10 μm. Representative image from two independent experiments. **e**, Left, immunofluorescent image of isolated crypt stained with pS6 antibody (red), Lgr5–eGFP (green) and DAPI (nuclei, blue). Scale bar, 10 μm. Representative of two independent experiments. Right, immunoblots of pS6 and pS6K from isolated crypts of young and old mice, and densitometric quantification (ratio to actin) ($n = 14$ mice per age group). An outlier (red) deviating >2 s.d. was removed from the analysis. **f**, Isolated Lgr5^{hi} stem cells (eGFP, green) from young and old mice, stained with pS6 antibody (red), phalloidin (F-actin, white), DAPI (nuclei, blue). Cells were distributed to pS6^{hi} (cells with higher than mean pS6 intensity) or pS6^{lo} (lower than mean pS6 intensity) categories. $n = 3$ independent experiments. **g**, Distribution of pS6 intensity in isolated Lgr5^{hi} cells from young and old mice ($n = 3$ mice per age group, number of cells analysed shown above the corresponding box and whisker plots). **h**, Mouse weights ($n = 25$ for young female, $n = 26$ for old female, $n = 20$ for young male, $n = 19$ for old male). Whiskers plotted according to Tukey's method. Y, mice between 3 and 9 months of age; O, mice over 24 months of age in all experiments. In box plots, unless otherwise indicated, the line represents median, the box shows interquartile range and whiskers represent the range. All other data are mean ± s.d.; two-tailed unpaired Student's *t*-test; exact *P* values shown in corresponding panels. For gel source data, see Supplementary Fig. 3.



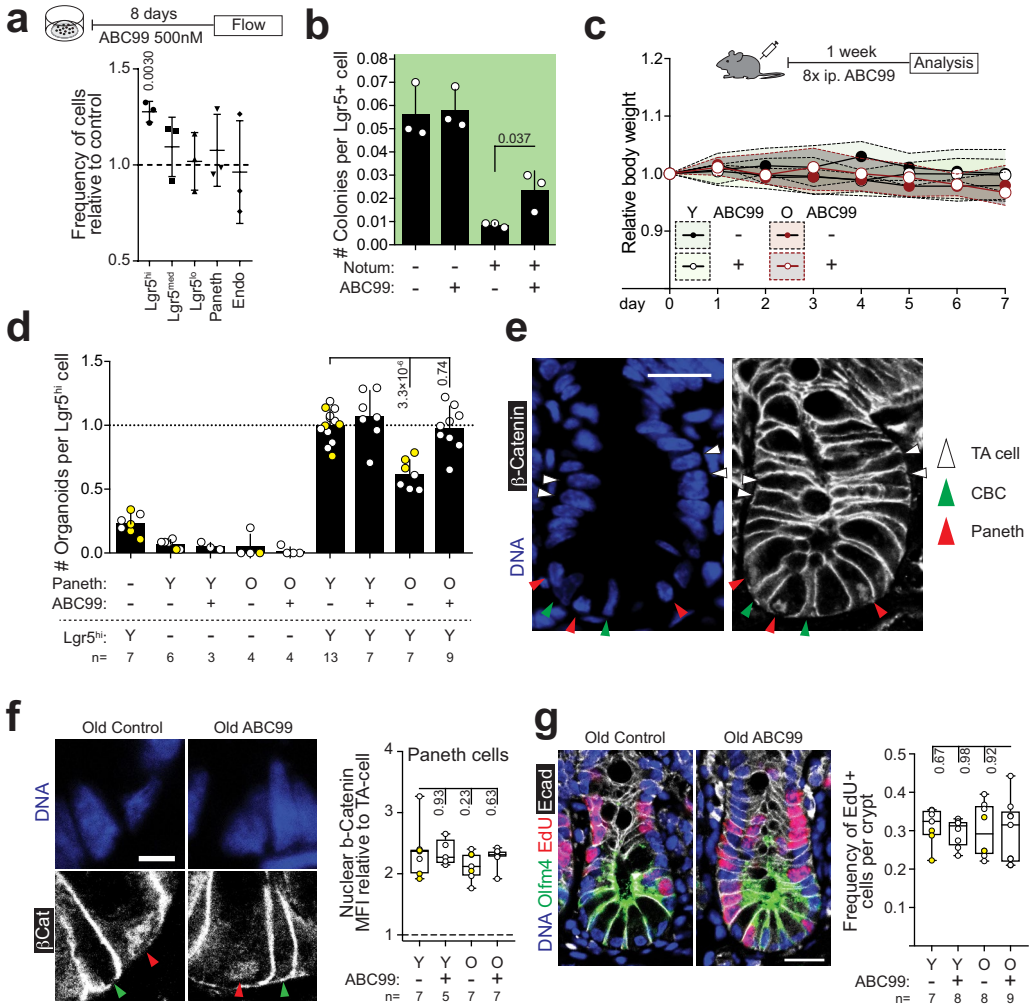
Extended Data Fig. 5 | Inhibiting mTORC1 activity in old mice restores intestinal regenerative capacity. a, Organoid-forming capacity and survival of subcultured intestinal organoids treated with rapamycin. Crypts were either treated continuously for four days (2 nM) or with a two-day pulse (2 nM pulse, 10 nM pulse) followed by two days in normal medium before subculturing and quantification ($n = 3$). **b**, Regenerative growth of organoids from young and old mice treated with 2 nM rapamycin for 2 days ex vivo. Crypt number was scored six to seven days after treatment from secondary subcultures (two days after passage) ($n = 5$ mice per age group). Student's paired *t*-test. Representative images are from subcultures on day 2. Scale bar, 100 μm. **c**, Weight of mice receiving daily injections of rapamycin (4 mg kg^{-1}) or vehicle ($n = 5$ mice per group). Daily data points represent median (circles) and interquartile range (dashed line). **d**, Immunoblots of pS6 from isolated crypts of vehicle (V)- or rapamycin (R)-treated young and old mice t ($n = 4$ mice per group). **e**, Organoid-forming capacity of isolated crypts from old mice treated with vehicle or rapamycin ($n = 4$ mice per group). **f**, Primary regenerative growth of organoids from old mice treated with vehicle or rapamycin ($n = 4$ mice per group). **g**, Organoid-forming capacity of young Lgr5^{hi} stem cells co-cultured with Paneth cells isolated from young or old mice treated with vehicle or rapamycin ($n = 4$ mice per group). Combinations compared to average of co-cultures with young vehicle- and old rapamycin-treated Paneth cells. **h**, Clonogenic growth of Lgr5^{hi} stem cells from young or old mice treated with vehicle or rapamycin ($n = 4$ mice per group); colonies quantified at day 7. **i**, RT–qPCR analysis of relative *Wnt2b*, *Wnt5a*, *Wnt4*, *Wnt3* and *Lgr5* expression from full jejunal samples of old mice treated with rapamycin. Values show fold change (expressed in log_2) in comparison to old vehicle treated (*n* values of mice analysed shown). Data are mean ± s.e.m. **j**, RT–qPCR analysis of relative *Notum* and *Bst1* expression from crypts of old mice treated with rapamycin. Values show fold change (expressed in log_2) in comparison to old vehicle-treated ($n = 3$ mice per group). Data are mean ± s.e.m. **k**, Immunoblots of pS6, S6 and H3 from isolated Epcam^{+} cells of wild type (Tsc1^{WT}) and Tsc1 knockout (Tsc1$^{\Delta}$) epithelium ($n = 3$ mice per group). **l**, Quantification of RNA-scope for *Notum* mRNA in wild type (Tsc1(WT)) and Tsc1 knockout (Tsc1(Δ)) ileal crypts ($n = 6$ mice for Tsc1(WT) and 5 mice for Tsc1(Δ)). An outlier (red) deviating >2 s.d was removed from the analysis. Representative images of crypts used in quantifications with *Notum* mRNA (brown) in Paneth cells (inset). **m**, Organoid-forming capacity of isolated crypts from Tsc1(WT and Tsc1(Δ) epithelium. Quantification was done on day 8. Y, mice between 3 and 9 months of age; O, mice over 24 months of age in all experiments. For gel source data, see Supplementary Fig. 3.



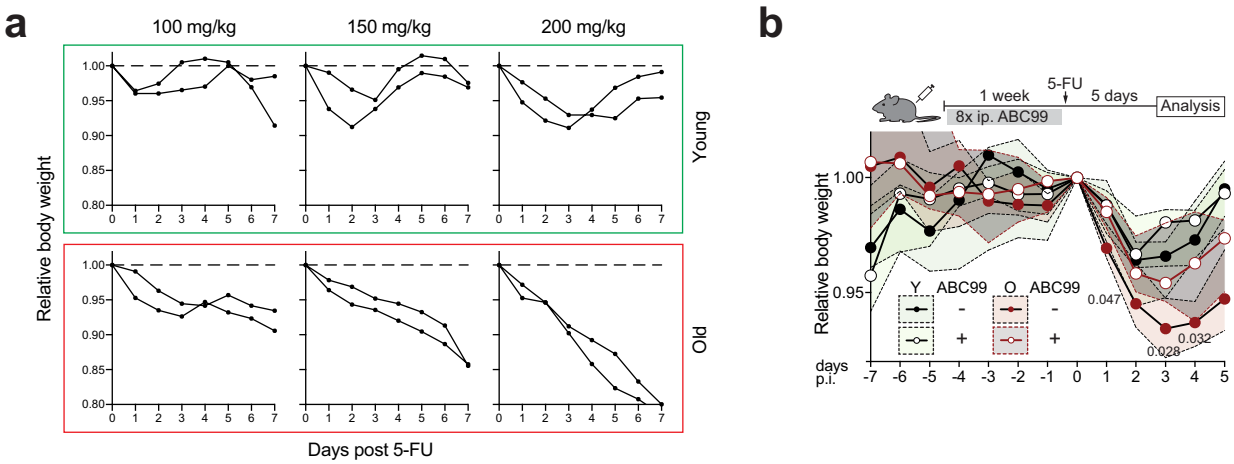
Extended Data Fig. 6 | Decreased PPAR activity in aged intestine. **a**, GSEA analysis for 'Biocarta PPARa' and 'PPARd' gene sets (for statistics, see 'RNA sequencing and data processing' in Methods). Nominal *P* value is shown ($n = 5$ mice in old, $n = 4$ mice in young). **b**, Schematic of the putative PPARα-binding site on the mouse and human *NOTUM* genes found with DECODE. Mouse sequence shown. Bottom, score for the discovered site using JASPAR matrix models for mouse PPAR-response element (PPRE). PPARG; RXRA motif was used. **c**, FACS analysis of cell populations in primary organoids treated for three days with DMSO, CHIR99021 or GW6471 ($n = 6$ mice for DMSO, $n = 5$ mice for CHIR99021 and GW6471). Ratios of $Lgr5^{hi}$ to Paneth cells, and $Lgr5^{hi}$ to $Lgr5^{lo}$ cells from the same analysis. Mean CD24 expression of live $Epcam^{+}$ cells are also shown. **d**, Representative images of mouse intestinal organoids treated for 4 days with DMSO, 5 μM GW6471 or 5 μM GW6471 + 100 ng ml^{-1} Wnt3A. Arrowheads indicate surviving, and red asterisks collapsed, organoids. Scale bar, 100 μm. Experiment was repeated four times with similar results. Unless otherwise indicated, data are mean ± s.d.; two-tailed unpaired Student's *t*-test; exact *P* values shown in corresponding panels.



Extended Data Fig. 7 | Notum regulates intestinal stem cell function. **a**, *Notum* gene targeting. Schematic represents sites of genome editing. Gene editing was confirmed by PCR with primers flanking the editing site (174 bp product for Notum KO1 and 294 bp product for Notum KO2) and hitting the edited site (84 bp product for Notum KO1 and 188 bp product for Notum KO2). Representative agarose gel images from two independent experiments with similar results are shown. **b**, Regenerative growth of Notum knockout organoids. De novo crypt domains were quantified two days after subculture ($n = 5$ repeated experiments with the same organoid lines). Representative images of organoids two days after subculture are shown. Scale bar, 100 μm. **c**, Schematic presenting in vivo competition assay of gene-edited organoid growth by orthotopic transplantation to immunodeficient *Rag2*$^{-/-}$ mice. Representative colonoscopy, necroscopy and histology images used for assay quantification ($n = 8$ mice transplanted). Scale bars, 1 mm for necroscopy and 200 μm for histology. **d**, Representative images of CRISPR-targeted young and old organoids two days after subculturing ($n = 4$ mice per group). Scale bar, 100 μm. **e**, Relative *Notum* expression in organoids with synergistic activator mediator complex (SAM) targeted to *Notum* promoter (dA Notum) grown for two days in ENR medium. Three independent experiments; relative to control (dA Tom). **f**, Quantification and representative images of day-5 colonies formed by isolated CD24^{med}SSC^{lo} cells from *Notum* activator (dA Notum) and control (dA Tom) organoids. Scale bar, 100 μm. $n = 4$ repeated experiments with the same organoid line. **g**, RT–qPCR analysis of relative *Axin2* and *Lgr5* expression in CD24^{med}SSC^{lo} cells sorted from *Notum* activator (dA Notum) organoids. Values show fold change (expressed in $\log_2$) in comparison to control (dATom) ($n = 3$ replicate wells per organoid line). In box plots, unless otherwise indicated, the line represents median, the box shows interquartile range and whiskers represent the range. All other data are mean ± s.e.m.; two-tailed unpaired Student's *t*-test; exact *P* values shown in corresponding panels. For gel source data see Supplementary Fig. 3.



Extended Data Fig. 8 | Notum inhibitor ABC99 prevents Wnt inactivation. a, Flow cytometry analysis of cell populations in primary organoids treated for eight days with 500 nM ABC99 ($n = 3$ mice) relative to DMSO control. Student's paired *t*-test. **b**, Clonogenic growth of Lgr5^{hi} stem cells on day 5 treated with or without 50 nM ABC99 and/or 500 ng ml^{-1} recombinant Notum (two independent experiments with similar results, one experiment with three replicate wells shown). **c**, Relative weight of mice treated with daily injections of ABC99 (10 mg per kg (body weight)) or control (vehicle or ABC101 10 mg per kg (body weight)) ($n = 10$ mice for young control and young ABC99, $n = 8$ mice for old control and $n = 9$ mice for old ABC99). Daily data points represent median (circles) and interquartile range (dashed line). **d**, Clonogenic growth of young Lgr5^{hi} stem cells co-cultured with young or old Paneth cells from mice treated with ABC99 or control (*n* values for analysed mice shown). Combinations compared to average of co-cultures with young control (−) and old ABC-treated (+) Paneth cells. Control mice received an equal amount of the inactive analogue ABC101 (yellow circles) or vehicle. **e**, Representative image of immunofluorescent staining of ileal crypts used for quantification of nuclear β-catenin (white) intensity. Paneth cells (red arrowheads) and CBCs (green arrowheads) were identified by cellular and nuclear (DAPI, blue) morphology. Their nuclear β-catenin levels were compared to transit-amplifying cells (white arrowheads). Scale bar, 20 μm. Experiment was repeated twice with a total of 26 mice all showing strongest nuclear β-catenin at the crypt bottom. **f**, Immunofluorescent staining of histological sections from old ileum. β-catenin (white), lysozyme (red) and DAPI (nuclei, blue). Scale bar, 10 μm. Quantification of relative nuclear β-catenin intensity of Paneth cells (red arrowhead) (*n* values for analysed mice shown). For quantification of CBCs (green arrowhead) see Fig. 3d. **g**, Immunofluorescent staining of histological sections from old ileum. Olfm4, green; EdU, red; DAPI (nuclei), blue. Scale bar, 20 μm. Quantification of EdU^{+} cellular frequencies within the crypt (*n* values for analysed mice shown). Y, mice between 3 and 9 months of age; O, mice over 24 months of age in all experiments. In box plots, unless otherwise indicated, the line represents median, the box shows interquartile range and whiskers represent the range. All other data are mean ± s.d.; two-tailed unpaired Student's *t*-test; exact *P* values shown in corresponding panels.



Extended Data Fig. 9 | Old intestine recovers poorly from 5-FU-induced damage. a, Body weights of young and old mice following single injection of 5-FU (100–200 mg kg^{-1}). Two mice per group, body weight relative to day of injection (day 0). **b**, Relative body weight of young and old mice treated for one week with or without ABC99 followed by single 5-FU (100 mg kg^{-1}) injection ($n = 8$ mice for young vehicle, old vehicle and young ABC99-treated, $n = 10$ mice for old ABC99-treated). Daily data points represent median (circles) and interquartile range (dashed line). Daily weight of old ABC99-treated mice were compared to old controls with two-tailed unpaired Student's *t*-test; exact *P* values shown under the corresponding daily weight. Young mice between 3 and 4 months of age; old mice over 20 months of age.

nature research

Corresponding author(s): Pekka Katajisto

Last updated by author(s): Apr 25, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	FACSDiva versions 7 and 8 Image J 64bit Fiji Leica Application Suite QuPath
Data analysis	Microsoft Excel 16.16.8 GraphPad Prism 8 FACSDiva versions 7 and 8 FlowJo v10 ImageJ 64bit cutadapt versio 1.8.1 STAR versiot 2.4.2a ja 2.6.1d DESeq2 versio 1.10.1 GenomicFeatures 1.22.13 GenomicAlignments 1.6.3

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

RNA sequencing data is publicly available through ArrayExpress (http://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-7916). All other data are available from the corresponding author upon reasonable request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was not predetermined with any statistical test. Sufficient sample sizes were estimated based on previous experiments with the same system.
Data exclusions	Outliers in Extended Data Fig. 4e and Extended Data Fig. 5l were removed from statistical analysis based on deviating >2 s.d. from the rest. Outliers are marked red in the corresponding figures. Cut-off of 2 s.d. was applied to all data of similar type, but exclusion criteria was not pre-set. No other data was excluded for other reasons than for being technically inadequate or unusable due to mistakes.
Replication	All experiments were successfully reproduced with biological replicates or independent technical replicates, except data in Extended Data Fig. 1o , where only three replicates from the same mouse were performed with similar results and in Extended Data Fig. 2d where KO validation was performed only once for the organoid line used in Extended Data Fig. 2e.
Randomization	Animals were allocated to experimental groups randomly, but no formal randomization were performed
Blinding	All histological analysis were performed by blinded investigator. In vitro quantifications were analyzed by blinded investigator when possible, but due to features of co-culture experiments this was not always possible. For animal work, investigators were not blinded due to evident external phenotype of aged mice.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	CD31-PE, Biolegend, Mec13.3, 1:500 CD45-PE, eBioscience, 30-F11, 1:500 Ter-119-PE, Biolegend, Ter119, 1:500 CD24-PacificBlue, Biolegend, M1/69, 1:500 Epcam-APC, eBioscience, G8.8, 1:500 Olfm4 (clone PP7), gift from CST in Fig 1d, 1:300 Olfm4, CST, 39141 in Fig 3f, 1:300

Lysozyme, DAKO, EC3.2.1.17, 1:750 (immunofluorescence) and 1:500 (cytochemistry)
pS6 (ser240/244), CST,5364, 1:1000
pS6 (ser235/236), CST, 4858, 1:500
S6, CST, 2217, 1:500
beta-Catenin, BD, 610153, 1:300
E-cadherin, BD, 610181, 1:500
Ki67, abcam, ab15580, 1:300
H3, CST, 4499, 1:1000
beta-Actin, CST, 4967, 1:2000
alpha-Tubulin, CST, 2144, 1:1000
pS6K, ImmunoWay, YP0886, 1:500
Muc 2, Santa Cruz, H-300, 1:50
anti-rabbit-Alexa-488, ThermoFischer, A11008, 1:500
anti-rabbit-Alexa-594, ThermoFischer, A11012, 1:500
anti-mouse-Alexa-647, ThermoFischer, A21203, 1:500
anti-rabbit-Alexa-633, ThermoFischer, A21071, 1:500
anti-rabbit-Alexa-647, ThermoFischer, A21244, 1:500
anti-mouse-Alexa-647, ThermoFischer, A21235, 1:500
anti-rabbit HRP, Sigma-Aldrich, A0545, 1:5000
anti-mouse HRP, CST, 7076, 1:1000

Validation

Only commercial antibodies were used that were validated by the vendor, data available on the manufacturer's website. For immunostainings, replicate samples stained with only secondary antibodies were used to determine specificity of the primary antibody. For pS6 (S240/244) antibodies, signal reduction was observed in samples treated with rapamycin, known inhibitor of mTORC1 activity both in immunoblots and in immunofluorescence. For pS6 (S235/236) increase in corresponding band intensity was observed after deletion of Tsc1, a known inhibitor of mTORC1 activity. For Lysozyme antibody, specific staining was observed in phenotypic Paneth cells (large granular cells at the bottom of crypts of the small intestine). Olfm4 antibodies stained only cells at the base of crypts but not phenotypic Paneth cells. Ki-67 antibody stained nuclei of epithelial cells that were in crypts, and not outside the crypt where proliferative capacity disappears. Ecadherin stained membranes of epithelial cells and has been used in the laboratory before with similar results. beta-Catenin antibody stained membranes of all epithelial cells and nuclei of cells at the crypt base where high Wnt-activity is known to exist. Muc-2 antibody stains large mucous producing cells and a mucus layer on top of the epithelium along the crypt to villus axis in the small and large intestine in histological samples.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	293fT cells were purchased from Thermo Fischer Scientific,R70007, lot# 1745311
Authentication	Cell lines were not authenticated
Mycoplasma contamination	Cell lines were regularly tested for mycoplasma contamination. All cell lines tested negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used (ICLAC Version 8.0)

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mouse (Mus musculus): Wild type and Lgr5-EGFP-IRES-CreERT2 and Rag2(-/-) mice were C57BL/6J background. Villin-CreERT2, Tsc1(fl/fl), Rosa26(LSL-ZsGreen), Rosa26(LSL-TdTomato), Rosa26(mT/mG) mice and Rosa26(LSL-Cas9EGFP) mice were mixed background. In all experiments, animals used were between 3 and 26 months of age. Age groups are stated in corresponding figure legends. Both female and male mice were used throughout the study.
Wild animals	Study did not involve wild animals
Field-collected samples	Study did not involve field-collected samples
Ethics oversight	Experiments using laboratory mice were approved and carried out in accordance with the guidelines of the Finnish national animal experimentation board and the Committee on Animal Care at MIT

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Human research participants

Policy information about studies involving human research participants

Population characteristics	For human colonic and ileal biopsies, patients between 32-81 years old, 12 female and 12 male, for Figure 1a and patients between 21-88 years, gender not know, for Extended Data Fig. 1h,j. undergoing routine colonoscopy. Exclusion criteria included any history of malignancy, chronic liver disease, history suggesting a malabsorption disorder, previous intestinal surgery, renal

	disease, bleeding disorder that would preclude biopsy, active infection, or systemic inflammatory disorder. Human jejunal samples (Extended Data Fig. 2g) were obtained from patients undergoing Roux en-Y gastric bypass surgery.
Recruitment	For colonoscopy biopsies, patient were recruited when they were attending routine colonoscopy at the clinic. For Roux-en-Y gastric bypass, patients were recruited as a part of a larger study months prior to the surgery. Recruitment unlikely caused bias relevant to the current study. Written and informed consent was obtained prior to enrolment.
Ethics oversight	The study regarding relevant samples was approved by the institutional review board of Massachusetts General Hospital (Boston, Massachusetts) and Helsinki University Hospital.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Mouse primary intestinal epithelial cells were isolated by EDTA treatment of minced tissue followed by gentle mechanical dissociation and enzymatic treatment with TrypLE Express enzyme to yield single cell suspension.
Instrument	FACSAriaII and FACSAriaIII Fusion (BD) were used to analyse and collect data.
Software	FACSDiva 7 and 8 were used to collect the data. FlowJo v10 was used for analyzing cellular frequencies.
Cell population abundance	Lgr5+ stem cell populations were ~99 % pure based on fluorescent microscopy of the post-sort fraction. Paneth cells were ~90% pure based on phase contrast microscopy (granulated morphology) and Lysozyme staining of the post-sort fraction. Culture of Paneth cells alone was performed in order to analyze the frequency of Paneth-ISC doublets.
Gating strategy	Single cells were gated by using FSC-A,FSC-W,SSC-A and SSC-W parameters as detailed in Supplementary data 1. Initially the right population was identified by overlaying Lgr5-EGFP on SSC-A vs FSC-A gate. Similar overlay were used to ensure that doublets were not included in the downstream gating. For fluorescent markers, positive populations were identified by comparing to unstained control sample from the same tissue.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

LETTER

https://doi.org/10.1038/s41586-019-1311-3

Distinct modes of mitochondrial metabolism uncouple T cell differentiation and function

Will Bailis[1,2,12], Justin A. Shyer[1,12], Jun Zhao[1,3,4], Juan Carlos Garcia Canaveras[5,6,7], Fatimah J. Al Khazal[8], Rihao Qu[1,3,4], Holly R. Steach[1], Piotr Bielecki[1], Omair Khan[1], Ruaidhri Jackson[1], Yuval Kluger[3,4,9], Louis J. Maher III[8], Joshua Rabinowitz[5,6,7], Joe Craft[1,10]* & Richard A. Flavell[1,11]*

Activated CD4 T cells proliferate rapidly and remodel epigenetically before exiting the cell cycle and engaging acquired effector functions. Metabolic reprogramming from the naive state is required throughout these phases of activation[1]. In CD4 T cells, T-cell-receptor ligation—along with co-stimulatory and cytokine signals—induces a glycolytic anabolic program that is required for biomass generation, rapid proliferation and effector function[2]. CD4 T cell differentiation (proliferation and epigenetic remodelling) and function are orchestrated coordinately by signal transduction and transcriptional remodelling. However, it remains unclear whether these processes are regulated independently of one another by cellular biochemical composition. Here we demonstrate that distinct modes of mitochondrial metabolism support differentiation and effector functions of mouse T helper 1 (T_H1) cells by biochemically uncoupling these two processes. We find that the tricarboxylic acid cycle is required for the terminal effector function of T_H1 cells through succinate dehydrogenase (complex II), but that the activity of succinate dehydrogenase suppresses T_H1 cell proliferation and histone acetylation. By contrast, we show that complex I of the electron transport chain, the malate–aspartate shuttle and mitochondrial citrate export are required to maintain synthesis of aspartate, which is necessary for the proliferation of T helper cells. Furthermore, we find that mitochondrial citrate export and the malate–aspartate shuttle promote histone acetylation, and specifically regulate the expression of genes involved in T cell activation. Combining genetic, pharmacological and metabolomics approaches, we demonstrate that the differentiation and terminal effector functions of T helper cells are biochemically uncoupled. These findings support a model in which the malate–aspartate shuttle, mitochondrial citrate export and complex I supply the substrates needed for proliferation and epigenetic remodelling early during T cell activation, whereas complex II consumes the substrates of these pathways, which antagonizes differentiation and enforces terminal effector function. Our data suggest that transcriptional programming acts together with a parallel biochemical network to enforce cell state.

T cells require mitochondrial metabolism as they exit from the naive cell state to become activated, and as they return to being resting memory cells; however, the role of mitochondrial metabolism in the differentiation and function of effector T cells is less well-understood[3–5]. Metabolite tracing studies have revealed that, whereas activated T cells use glutamine for the anaplerosis of α-ketoglutarate, activated cells decrease the rate of pyruvate entry into the mitochondria in favour of lactate fermentation[5,6]. Despite the decreased utilization of glucose-derived carbon for mitochondrial metabolism, the tricarboxylic acid (TCA) cycle has previously been shown to contribute to IFNγ production by increasing cytosolic acetyl-CoA pools via mitochondrial citrate export[7]. Additionally, the TCA cycle can contribute to the electron transport chain (ETC) by generating NADH and succinate to fuel complex I and complex II, respectively. However, the role of the ETC in the later stages of T cell activation is poorly characterized. To test the contribution of the TCA cycle to the function of effector T cells, we treated cells cultured in T_H1 conditions with the TCA-cycle inhibitor sodium fluoroacetate[8]. We titrated sodium fluoroacetate or the glycolysis inhibitor 2-deoxy-D-glucose (2DG; an inhibitor of T_H1 cell activation, used as a positive control) at day 1 of T cell culture, and assayed cell proliferation at day 3 or the expression of the *Ifng*-Katushka reporter at day 5. Although 2DG was a more-potent inhibitor than sodium fluoroacetate at lower doses, both inhibitors impaired *Ifng* transcription (Fig. 1a) and T cell proliferation (Fig. 1b) in a dose-dependent manner, which suggests that the activity of TCA-cycle enzymes is required for optimal T_H1 cell activation.

To evaluate which processes downstream of the TCA cycle contribute to the role of the TCA cycle in T-helper-cell proliferation and function, we treated T_H1 cells with inhibitors of the ETC overnight on day 2 (to evaluate proliferation) or overnight on day 4 (to evaluate cytokine production), and analysed cells the following day. Unlike impairing glycolysis with 2DG or the TCA cycle with sodium fluoroacetate, which resulted in a block of both proliferation and function, we observed a dichotomy in the role of the ETC in supporting each of these processes. Although the inhibition of complex II did not impair proliferation, blocking complex I and complex III resulted in a decrease in the number of divided cells; treatment with oligomycin displayed a modest but significant effect (Fig. 1c). Importantly, viability was not affected upon acute inhibition of ETC complexes (Extended Data Fig. 1a). Consistent with this observation, treatment with rotenone or antimycin A on day 2 resulted in cell-cycle arrest at the G2 or M phase, whereas treatment with dimethyl malonate (DMM) or oligomycin did not alter cell-cycle status (Extended Data Fig. 1b). Similar to cells cultured in T_H1 conditions, cells cultured in T_H2 or T_H17 conditions displayed defects in proliferation and an altered cell cycle when treated with rotenone (Extended Data Fig. 2a, b, e, f), which suggests that complex I supports cell division regardless of the cytokine environment.

Further illustrating distinct roles for complex I and complex II in T-helper-cell proliferation and function, we observed that the ATP citrate lysase (ACLY) inhibitor BMS-303141 significantly decreased IFNγ production, consistent with previous work[7], whereas the effect of inhibition of complex I or ATP synthase with rotenone or oligomycin, respectively, was not significant. By contrast, impairing complex II activity with DMM, or complex III activity with antimycin A, significantly reduced IFNγ production to levels below those observed with BMS-303141 (Fig. 1d). Together, these observations suggest that the

[1]Department of Immunobiology, Yale School of Medicine, New Haven, CT, USA. [2]Department of Pathology, Children's Hospital of Philadelphia, Philadelphia, PA, USA. [3]Department of Pathology, Yale School of Medicine, New Haven, CT, USA. [4]Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA. [5]Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ, USA. [6]Department of Chemistry, Princeton University, Princeton, NJ, USA. [7]Diabetes Research Center, University of Pennsylvania, Philadelphia, PA, USA. [8]Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine and Science, Rochester, MN, USA. [9]Applied Mathematics Program, Yale University, New Haven, CT, USA. [10]Department of Internal Medicine (Rheumatology), Yale School of Medicine, New Haven, CT, USA. [11]Howard Hughes Medical Institute, Chevy Chase, MD, USA. [12]These authors contributed equally: Will Bailis, Justin A. Shyer. *e-mail: joseph.craft@yale.edu; richard.flavell@yale.edu

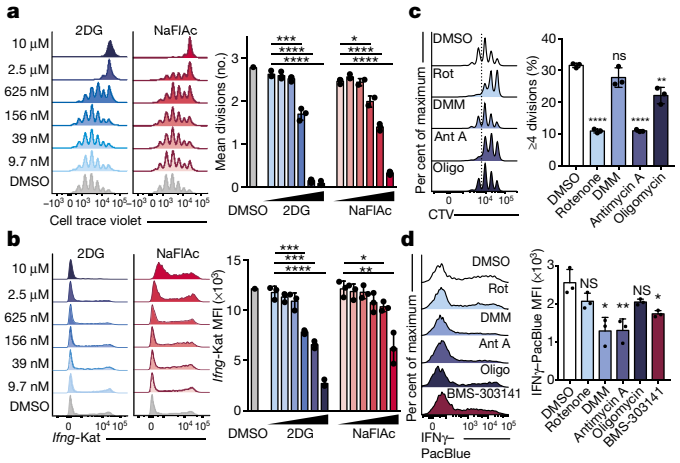


Fig. 1 | The TCA cycle supports proliferation and function of T helper cells through distinct mechanisms. a, **b**, Mean divisions at day 3 (**a**) and *Ifng*-Katushka (*Ifng*-Kat) reporter expression after restimulation with phorbol myristate acetate (PMA) and ionomycin at day 5 (**b**) of CD4 T cells cultured in T_H1 conditions with serially diluted 2DG ($n = 3$) or sodium fluoroacetate (NaFlAc) ($n = 2$ or 3). MFI, mean fluorescence intensity. **c**, **d**, Proliferation after overnight treatment on day 2 (**c**) and intracellular IFNγ protein expression after overnight treatment on day 4 (**d**) of wild-type CD4 T cells cultured in T_H1 conditions with dimethylsulfoxide (DMSO), rotenone (rot), DMM, antimycin A (ant A), oligomycin (oligo) or BMS-303141 ($n = 3$). *n*, number of technical replicates. Representative plots and a graph summarizing the results of at least two independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$. ns, not significant.

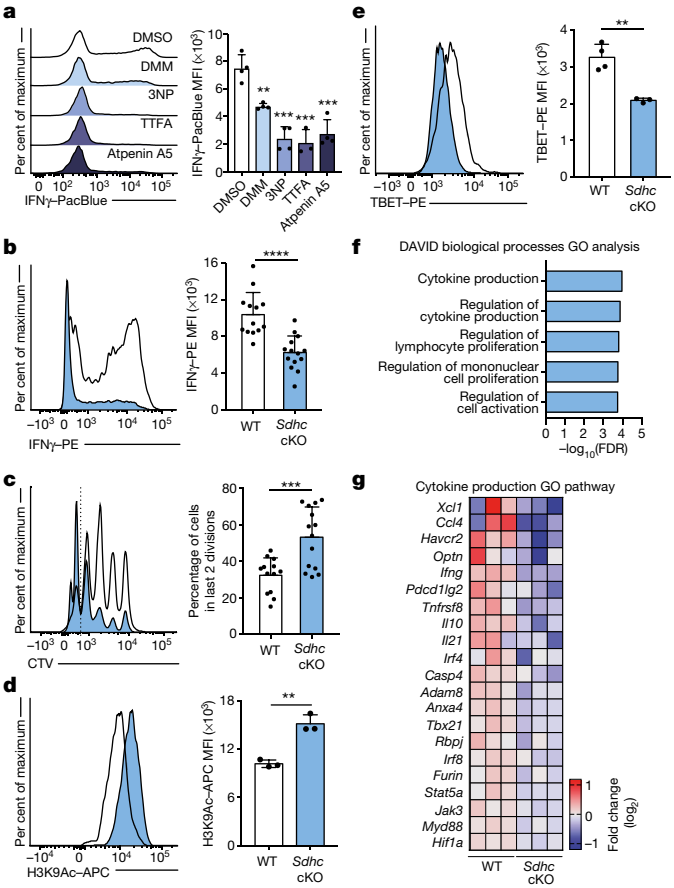


Fig. 2 | Complex II uncouples differentiation and effector function of T_H1 cells. a, Intracellular IFNγ protein expression in PMA and ionomycin-restimulated wild-type CD4 T cells cultured in T_H1 conditions at day 5 after overnight treatment with DMSO, DMM (10 mM), 3NP (1 mM), TTFA (100 μM) or atpenin A5 (1 μM) ($n = 3$). **b**, **c**, Intracellular IFNγ protein expression (**b**) and proliferation of CD4 T cells (**c**) from doxycycline-treated *Sdhc* cKO or wild-type (WT) mice cultured in T_H1 conditions at day 5. Data combined from 5 independent experiments: wild type, $n = 13$; *Sdhc* cKO, $n = 14$ biological replicates. Two-tailed *t*-test. **d**, Total cellular H3K9 acetylation (H3K9Ac) of wild-type and *Sdhc* cKO cells cultured in T_H1 conditions at day 3 ($n = 3$). Two-sided *t*-test. **e**, TBET protein expression of wild-type ($n = 4$) and *Sdhc* cKO ($n = 3$) cells cultured in T_H1 conditions at day 5. Two-sided *t*-test. **f**, DAVID Gene Ontology (GO) pathway analysis of genes that are downregulated in cKO mice compared to wild-type controls. $P < 0.05$. **g**, Heat map of gene expression from RNA-seq results for the cytokine production GO pathway. *n*, number of technical replicates, except where noted otherwise. Representative plots and a graph summarizing the results of at least two independent experiments are shown, except where noted otherwise. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

TCA cycle supports T_H1 function by enabling cytosolic acetyl-CoA production and by fuelling a succinate-dehydrogenase (SDH)-driven ETC. This role for the ETC was specific to the T-helper-cell cytokine culture conditions to which the cells were exposed during activation. Unlike T_H1 cells, inhibiting the ETC had a minimal effect on function of T_H2 effector cells; inhibition of complex I or complex III resulted in a slight, but significant, increase in IL-4 reporter activity (Extended Data Fig. 2c). By contrast, T_H17 cells displayed sensitivity to inhibition of both complex I and complex II (Extended Data Fig. 2d). These data indicate that the ETC has program-specific roles in regulating the effector functions of T helper cells.

To corroborate the effects of DMM on the function of T_H1 cells, we tested the capacity of three additional inhibitors of complex II—thenoyltrifluoroacetone (TTFA), 3-nitropropionic acid (3NP) and atpenin A5—to inhibit IFNγ production in T_H1 cells. Each drug impaired complex II activity, as assayed by cellular succinate accumulation (Extended Data Fig. 3a). Consistent with our results for DMM treatment, T_H1 cells treated with 3NP, TTFA or atpenin A5 produced significantly less IFNγ than control cells (Fig. 2a). In keeping with a role for the TCA cycle and complex II in promoting T_H1 cell function, cells cultured overnight with a membrane-permeable form of succinate (diethyl succinate) produced more IFNγ (Extended Data Fig. 3b). To genetically test the requirement of complex II activity in T_H1 cells, we generated a retroviral single-guide (sg)RNA expression vector (which we named MG-Guide) that is compatible with transduction of mouse T cells (Extended Data Fig. 4a, b). To validate the system, we transduced CD4 T cells with sgRNA and observed a rapid loss of protein expression when using sgRNAs that targeted *Tbx21* or *Il12rb1*, genes that are essential for T_H1-cell cytokine production; this loss led to a decrease in capacity for IFNγ production (Extended Data Fig. 4c–f, Supplementary Table 1). Transduction of T_H1 cells with a sgRNA targeting *Sdha*, which encodes the catalytic subunit of complex II, impaired capacity for IFNγ production (Extended Data Fig. 3c). To provide further genetic evidence that complex II activity is required for the function of T_H1 cells, we tested the requirement for *Sdhc*, which encodes an essential subunit of complex II. We cultured CD4 T cells isolated from $Sdhc^{fl/fl}$ *TetO-cre*$^{-/+}$ $Rosa26^{rtTA/+}$ (hereafter, *Sdhc* conditional knockout (cKO)) or $Sdhc^{+/+}$ *TetO-cre*$^{-/+}$ $Rosa26^{rtTA/+}$ control (hereafter, wild-type) mice that had been treated in vivo with doxycycline for ten days in T_H1 conditions. Unbiased mass-spectrometry analysis of metabolites in wild-type and *Sdhc* cKO T_H1 cells revealed that *Sdhc* cKO cells had increased levels of cellular succinate and α-ketoglutarate, which confirms the loss of SDH activity (Extended Data Fig. 3d, e). Consistent with our drug and sgRNA studies, *Sdhc* cKO cells produced significantly less IFNγ at day 5 post-activation (Fig. 2b). However, *Sdhc* cKO T_H1 cells proliferated significantly more than wild-type controls, which suggests that proliferation and effector function

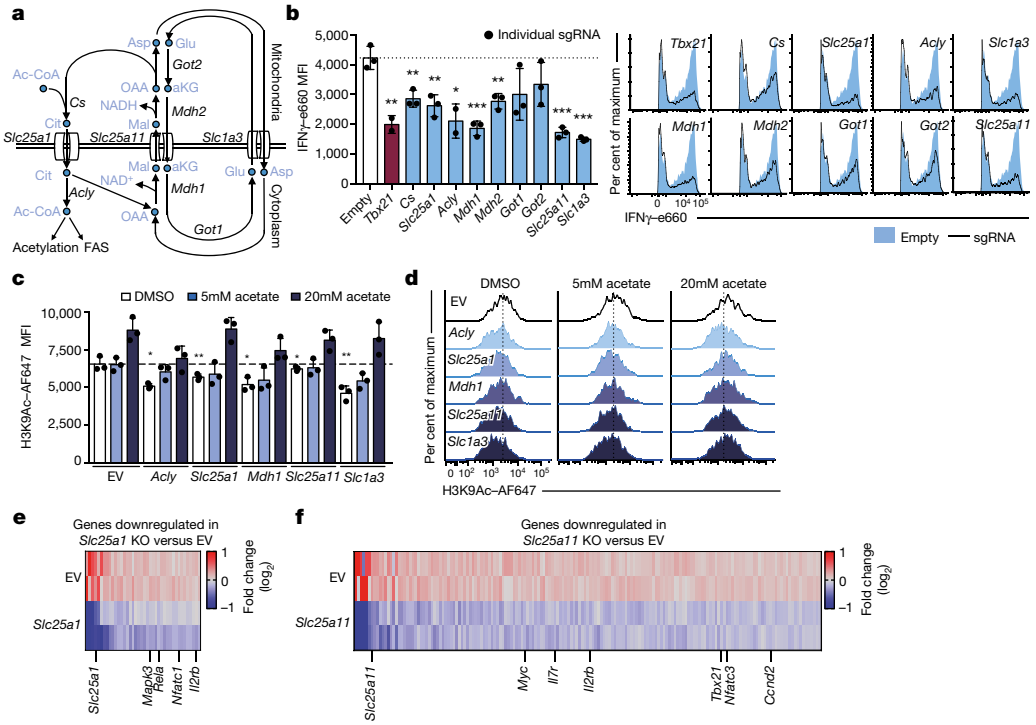


Fig. 3 | The malate–aspartate shuttle and mitochondrial citrate export are required for histone acetylation and proliferation in differentiating T_H1 cells. a, Schematic of the malate–aspartate shuttle and mitochondrial citrate export. aKG, α-ketoglutarate; Asp, aspartate; Cit, citrate; Glu, glutamate; Mal, malate; OAA, oxaloacetate. **b**, Intracellular IFNγ protein expression in Cas9-expressing CD4 T cells, transduced with sgRNAs targeting the indicated enzymes and transporters, cultured in T_H1 conditions after restimulation at day 5. Graphs show individual sgRNAs for each gene as well as the average for all three sgRNAs ($n = 2$ or 3 biological replicates). **c**, **d**, Total cellular H3K9 acetylation at day 4 of Cas9-expressing CD4 T cells transduced with sgRNAs against the indicated enzymes and transporters, in the absence or presence of 5 nM or 20 nM exogenous acetate added 1 day after transduction, cultured in T_H1 conditions ($n = 3$ technical replicates). **e**, **f**, Heat map summarizing downregulated genes determined by RNA-seq for cells expressing *Slc25a1*-targeting sgRNA (**e**) or *Slc25a11*-targeting sgRNA (**f**). $P < 0.05$. EV, empty vector; KO, knockout. Representative plots and a graph summarizing the results of at least two independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-sided *t*-test used to determine significance. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$.

are processes that are uncoupled by complex II activity (Fig. 2c). To test whether processes in addition to proliferation that are involved in T-helper-cell differentiation were affected, we assayed the effect of SDH deficiency on histone acetylation. We found that *Sdhc* cKO cells exhibited increased H3K9 acetylation, and that DMM treatment as well as delivery of *Sdha*-targeting sgRNA increased H3K9 and H3K27 acetylation; this suggests that complex II antagonizes T-helper-cell differentiation by negatively regulating both proliferation and histone acetylation (Fig. 2d, Extended Data Fig. 5a–c).

To test the role of complex II in promoting other aspects of the functional program of T_H1 cells, we evaluated TBET protein expression in *Sdhc* cKO and wild-type cells on day 5 after activation. Consistent with defects in IFNγ production, T_H1 cells from *Sdhc* cKO mice had reduced levels of expression of TBET protein (Fig. 2e). To further investigate a role for complex II in supporting the functional program of T_H1 cells, we performed RNA sequencing (RNA-seq) on effector T_H1 cells from *Sdhc* cKO and wild-type mice at day 5 after activation. Consistent with a decrease in TBET expression, T_H1 cells from mice deficient in complex II exhibited significantly decreased expression of genes that are key to the T_H1 cell program and genes that are important during T-helper-cell activation. Notably, DAVID (Database for Annotation, Visualization and Integrated Discovery) Gene Ontology pathway analysis indicated 'cytokine production' and 'regulation of lymphocyte proliferation' as the most-dysregulated pathways (Fig. 2f, g, Extended Data Fig. 5d, e, Supplementary Table 2). These data indicate that SDH activity is a primary mechanism through which mitochondrial metabolism supports the functional programming of T_H1 cells.

We next sought to investigate which aspects of mitochondrial metabolism are antagonized by SDH to constrain proliferation. The consumption of α-ketoglutarate is known to modulate the activity of mitochondrial shuttling systems that are required to maintain the cellular redox balance and the production of key cytosolic metabolites[9–11]. The malate–aspartate shuttle and mitochondrial citrate export are two such systems; they regulate the oxidation state of nicotinamide adenine dinucleotides (NAD) in the mitochondria and the transport of acetyl-CoA from the mitochondria to the cytosol, respectively. On the basis of our data that *Sdhc* cKO T_H1 cells exhibit increased proliferation (Fig. 2c) and increased cellular α-ketoglutarate levels (Extended Data Fig. 3e), we hypothesized that these mitochondrial transport systems promote the early stages of T_H1 cell proliferation.

To test the requirement of these transport systems for T_H1 cell activation, we designed three sgRNAs per gene of interest and conducted individual sgRNA knockout experiments using MG-Guide, measuring IFNγ protein (Fig. 3a). We found that, compared to cells transduced with an empty MG-Guide vector, cells that express sgRNAs that target *Mdh1*, *Mdh2*, *Slc25a11* or *Slc1a3* produced less IFNγ protein—comparable to the levels observed with sgRNAs that target the positive-control *Tbx21* gene—as did two of the three sgRNAs designed to target *Got1* and *Got2*, which suggests that the malate–aspartate shuttle is critical during T_H1 cell activation (Fig. 3b). In addition, we observed defective IFNγ production in T_H1 cells that express sgRNA against *Cs*, *Slc25a1* and *Acly*, which indicates that citrate synthesis and export for cytosolic acetyl-CoA production are also required (Fig. 3b).

Previous reports have suggested that ACLY activity is required for T_H1-cell histone acetylation, and the ETC has previously been shown to support epigenetic remodelling[7,12]. To test the role of both shuttle systems during T_H1-cell epigenetic remodelling, we evaluated total cellular H3K9 and H3K27 acetylation. We found that impairing *Acly*, *Slc25a1*, *Mdh1*, *Slc25a11* and *Slc1a3* results in decreased H3K9 acetylation, and that acetate supplementation could compensate for these

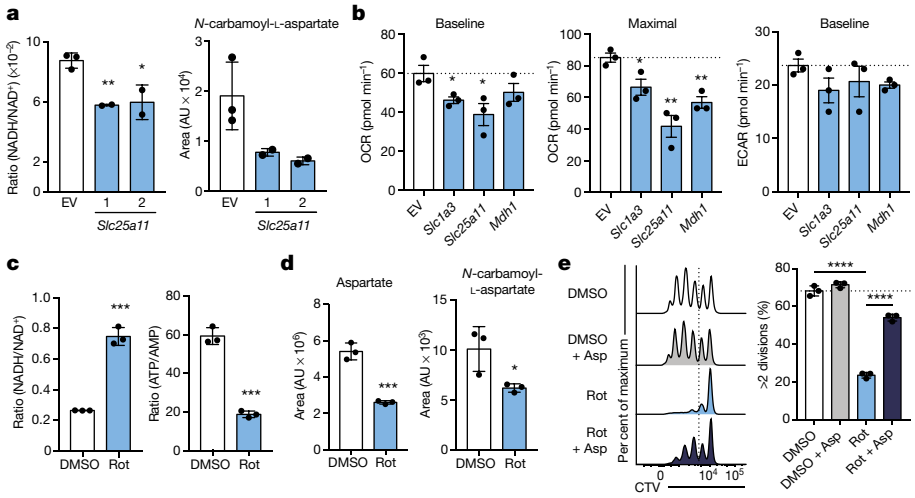


Fig. 4 | The malate–aspartate shuttle promotes complex I activity, which is required for aspartate synthesis and T-helper-cell proliferation. **a**, Cellular NADH/NAD$^+$ ratio and *N*-carbamoyl-L-aspartate measured by liquid chromatography–mass spectrometry analysis in Cas9-expressing CD4 T cells transduced with sgRNA targeting *Scl25a11*, and cultured in T_H1 conditions as described in Methods ($n = 2$ biological replicates, $n = 2$ technical replicates). AU, arbitrary units. **b**, Baseline oxygen consumption rate (OCR), maximal OCR and baseline extracellular acidification rate (ECAR) of Cas9-expressing CD4 T cells transduced with sgRNAs targeting the indicated enzymes and transporters, cultured in T_H1 conditions at day 4 ($n = 3$ biological replicates). **c**, **d**, Cellular NADH/NAD$^+$ and ATP/AMP ratios (**c**) and aspartate and *N*-carbamoyl-L-aspartate (**d**) measured by liquid chromatography–mass spectrometry analysis in wild-type CD4 T cells cultured in T_H1 conditions, and treated with DMSO or rotenone for 4 h on day 4 ($n = 3$ technical replicates). **e**, Proliferation measured at day 3 of wild-type CD4 T cells cultured in T_H1 conditions and treated on day 2 with DMSO (clear and grey bar) or rotenone (blue bars) ± 20 mM aspartate ($n = 3$ technical replicates). Representative plots and a graph summarizing the results of at least two independent experiments are shown. Mean and s.d. are presented on summarized plots and unpaired, two-sided *t*-test used to determine significance. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

defects (Fig. 3c, d). By contrast, H3K27 acetylation was largely unaffected by targeting these genes (with the exception of *Slc25a1*); however, the addition of acetate resulted in increased H3K27 acetylation regardless of the condition (Extended Data Fig. 6a). This effect of acetate on histone acetylation is largely explained by an increase in total H3 content, whereas the effect of the sgRNA on acetylation is only partially explained by changes in total histone mass (Extended Data Fig. 6b–d).

To evaluate the transcriptional effects of deficiency in the malate–aspartate shuttle, we performed RNA-seq at day 5 after activation on T_H1 cells that express sgRNA against *Slc25a1* or *Slc25a11*. Consistent with a role for the shuttles in promoting T_H1 cell differentiation, we observed decreased expression of genes with known roles in T cell activation and T_H1 cell programming. Targeting either of the transporters led to impaired expression of *Il2rb*, whereas loss of *Slc25a1* affected key T-cell-activation genes (such as *Nfatc1*, *Rela* and *Mapk3*) and disruption of *Slc25a11* resulted in the loss in expression of genes including *Tbx21*, *Nfatc3*, *Ccnd2* and *Myc* (Fig. 3e, f, Extended Data Fig. 6e, f, Supplementary Tables 3, 4).

Given the importance of *Il2rb, Myc* and *Ccnd2* in T-helper-cell division, we next evaluated the role of the shuttles in regulating T-helper-cell proliferation. To test this, we evaluated cell division in cells cultured in T_H1 conditions that express sgRNAs targeting *Acly*, *Slc25a1*, *Mdh1*, *Slc25a11* or *Slc1a3*. Relative to controls, targeting any of these genes resulted in modestly—but significantly—decreased proliferation (Extended Data Fig. 7). Collectively, these data demonstrate that the malate–aspartate shuttle and mitochondrial citrate export are required for T_H1 cell proliferation and transcriptional remodelling.

To investigate the biochemical mechanism that might explain these observations, we performed mass-spectrometry analysis of T cells transduced with guides targeting either *Slc25a1* or *Slc25a11* sgRNA. As expected, we found that disrupting citrate transport results in decreased levels of cellular acetyl-CoA (Extended Data Fig. 8a–c). Unexpectedly, targeting *Slc25a11* resulted in a decreased cellular NADH/NAD$^+$ ratio, which suggests that the activity of complex I is a primary mechanism by which cellular NADH/NAD$^+$ is regulated in activated T_H1 cells (Fig. 4a, Extended Data Fig. 8d, e). Moreover, targeting either shuttle system resulted in diminished levels of intermediates of the pentose phosphate pathway and of *N*-carbamoyl-L-aspartate, an essential precursor molecule for nucleotide synthesis (Fig. 4a, Extended Data Figs. 8b, c, 9a, b). Consistent with a role for the shuttling systems in providing mitochondrial NADH for the ETC, Seahorse analysis demonstrated that rates of basal and maximal oxygen consumption were impaired upon expression of sgRNAs targeting either *Mdh1*, *Slc25a11* or *Slc1a3* (Fig. 4b). This was not substantially compensated for by increased glycolysis, as the extracellular acidification rate was minimally affected (Fig. 4b).

Having observed that complex I supports early T-helper-cell proliferation and that the malate–aspartate shuttle fuels complex I (Fig. 1c), we next sought to examine the biochemical mechanism by which complex I promotes proliferation by performing mass-spectrometric analysis on rotenone-treated cells. As expected, inhibiting complex I increased the NADH/NAD$^+$ ratio and decreased the ATP/AMP ratio (Fig. 4c, Extended Data Fig. 9a, b). Rotenone treatment also led to decreased pools of cellular aspartate and *N*-carbamoyl-L-aspartate in these cells, similar to previous observations in cancer-cell lines[13,14] (Fig. 4d). To test whether this aspartate synthesis deficiency contributed to the proliferative defects of rotenone-treated cells, we supplemented rotenone-treated cells with aspartate and evaluated cell division and the cell cycle. Aspartate supplementation resulted in a significant recovery of cell proliferation, and a partial release from the arrest at the G2 or M phase following rotenone treatment (Fig. 4e, Extended Data Fig. 9c). These data demonstrate that the regulation of complex I by mitochondrial shuttling systems determines the cellular redox balance and the cytosolic aspartate availability that is required for T cell proliferation.

Using approaches that combine network-level genetic interrogation of metabolic pathways, pharmacology, transcriptomics and metabolomics, we demonstrate how T_H1 cells meet the distinct metabolic demands of differentiation and function during the course of activation. To generate the substrates needed for proliferation and epigenetic remodelling, early activated T helper cells fuel complex I through the malate–aspartate shuttle and mitochondrial citrate export. Unlike the carbon-neutral malate–aspartate shuttle (which exchanges malate for α-ketoglutarate), complex II moves carbon forward in the TCA cycle; this restricts processes that support differentiation and promotes the late-stage effector function of T_H1 cells, which permits cells to exit the

cell cycle and adopt their terminal program (Extended Data Fig. 10). These findings illustrate how differentiation and terminal effector function—previously understood to be concordantly regulated by signal transduction—are controlled by distinct metabolic modules, which elucidates how cell programming is governed by parallel transcriptional and biochemical networks.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1311-3.

Received: 8 November 2017; Accepted: 22 May 2019;
Published online 19 June 2019.

1. Buck, M. D., Sowell, R. T., Kaech, S. M. & Pearce, E. L. Metabolic instruction of immunity. *Cell* **169**, 570–586 (2017).
2. Buck, M. D., O'Sullivan, D. & Pearce, E. L. T cell metabolism drives immunity. *J. Exp. Med.* **212**, 1345–1360 (2015).
3. Klein Geltink, R. I. et al. Mitochondrial priming by CD28. *Cell* **171**, 385–397.e11 (2017).
4. Buck, M. D. et al. Mitochondrial dynamics controls T cell fate through metabolic programming. *Cell* **166**, 63–76 (2016).
5. Chang, C.-H. et al. Posttranscriptional control of T cell effector function by aerobic glycolysis. *Cell* **153**, 1239–1251 (2013).
6. Wang, R. et al. The transcription factor Myc controls metabolic reprogramming upon T lymphocyte activation. *Immunity* **35**, 871–882 (2011).
7. Peng, M. et al. Aerobic glycolysis promotes T helper 1 cell differentiation through an epigenetic mechanism. *Science* **354**, 481–484 (2016).
8. Peters, R. *Biochemical Lesions and Lethal Synthesis* (Pergamon, 1963).
9. Contreras, L. & Satrústegui, J. Calcium signaling in brain mitochondria: interplay of malate aspartate NADH shuttle and calcium uniporter/mitochondrial dehydrogenase pathways. *J. Biol. Chem.* **284**, 7091–7099 (2009).
10. Safer, B. The metabolic significance of the malate-aspartate cycle in heart. *Circ. Res.* **37**, 527–533 (1975).
11. LaNoue, K. F. & Williamson, J. R. Interrelationships between malate-aspartate shuttle and citric acid cycle in rat heart mitochondria. *Metabolism* **20**, 119–140 (1971).
12. Wellen, K. E. et al. ATP-citrate lyase links cellular metabolism to histone acetylation. *Science* **324**, 1076–1080 (2009).
13. Birsoy, K. et al. An essential role of the mitochondrial electron transport chain in cell proliferation is to enable aspartate synthesis. *Cell* **162**, 540–551 (2015).
14. Sullivan, L. B. et al. Supporting aspartate biosynthesis is an essential function of respiration in proliferating cells. *Cell* **162**, 552–563 (2015).

METHODS

T cell assays and sgRNA delivery. CD4 T cells were isolated from constitutive Cas9-expressing (Cas9tg) B6 mice[15], stimulated with anti-CD3 and anti-CD28 coated beads (Miltenyi T Cell Activation/Expansion Kit, mouse), and cultured in assay-determined T_H1 conditions (5 ng/ml IL-2, 2 ng/ml IL-12 and 10 μg/ml anti-IL-4). On day 1 post-activation, T cells were transduced with MG-Guide retrovirus using spin transduction at 1,200*g* for 90 min at 37 °C. IFNγ cytokine was measured by adding brefeldin A, 1 h after the addition of PMA (20 ng/ml) and ionomycin (20 ng/ml); 4 h after restimulation, cells were fixed, stained with anti-CD4 (Biolegend), anti-GFP (Millipore) and anti-IFNγ (Biolegend), and analysed by flow cytometry. To assay for *Ifng*-Katushka, IL-4–GFP, and IL-17–GFP expression, T cells from *Ifng*-Katushka[16], 4GET (Jackson Labs, 004190) and IL-17–GFP (Jackson Labs, 018472) reporter mice were activated with PMA and ionomycin for four hours, stained with anti-CD4 and then analysed by flow cytometry for reporter activity in GFP^+ cells. Cell division was measured by labelling cells with CellTrace Violet (Thermo) before activation, and evaluated for proliferation at day 3 after activation; where indicated, inhibitors and metabolites were added to the medium overnight on day 2 after activation. Cell-cycle status was determined by intracellular flow cytometry analysis of Ki67 and DAPI, at day 3 after activation; where indicated, inhibitors and metabolites were added to the medium overnight on day 2 after activation. Mitochondrial reactive oxygen species was measured by flow cytometry in CD4 T cells by staining cells with MitoSOX Red mitochondrial superoxide indicator (Thermo) and anti-CD4 for 30 min at 37 °C in the presence of the indicated inhibitors. For all experiments using inhibitors or metabolite supplementation, the following doses were used: 1 μM rotenone (Sigma), 10 mM DMM (Sigma), 1 mM 3NP (Sigma), 100 μM TTFA (Sigma), 1 μM atpenin A5 (Cayman Chemical), 1 μM antimycin A (Sigma), 1 μM oligomycin (Sigma), 5mM diethyl succinate (Sigma) or 20mM aspartate. All mice required for this study were housed and maintained under specific-pathogen-free conditions in the animal facility of the Yale University School of Medicine, and all corresponding animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Yale University. This study was conducted in compliance with all relevant ethical regulations. All cells used for experimentation were collected from male and female mice at 6–8 weeks of age.

MG-Guide vector generation, sgRNA cloning and retroviral production. MG-Guide was generated by removing the IRES element from MIGR1 (Addgene) by EcoRI and NotI digestion, and adding the human U6 promoter and SV40 promoter from pMKO-GFP (Addgene) by infusion assembly (Clonetech). To add the sgRNA cloning site, the vector was digested with AgeI and EcoRI and combined by infusion assembly with an IDT Gene Block containing two BbsI restriction sites upstream of a scaffold RNA sequence and a U6 stop. To clone individual sgRNAs, MG-Guide was digested with BbsI and pairs of oligonucleotides (Sigma) with complimentary overhangs were annealed and ligated into the vector. For retroviral production, 1 μg of MG-Guide plasmid and 0.5 μg of EcoHelper plasmid were transfected into 5×10^5 HEK293T cells (source ATCC, identity unconfirmed, not tested for mycoplasma) in a 6-well plate using X-tremeGENE 9 DNA Transfection Reagent (Roche) overnight. The medium was then replaced, and virus was collected 24 h later. Isolated CD4 T cells (1×10^6) were stimulated overnight, and spin-transduced in the viral preparation with 1 μg/ml polybrene at 1,200*g* for 90 min at 37 °C.

RNA-seq analysis. Raw reads from RNA-seq were aligned to the mouse genome mm10 with STAR 2.7.0[17], and gene-expression levels were measured by HTSeq 0.11.1[18]. Subsequently, differential expression analysis between different groups was performed with DESeq2[19].

Seahorse analysis. Analysis was performed on cells at day 3, day 4 and day 5 after activation. Cells were washed three times in complete Seahorse medium (Seahorse Bioscience) with 10 mM glucose, 1 mM sodium pyruvate and 2 mM glutamine. Cells were plated at 4×10^4 cells per well in a 96-well Seahorse assay plate, pretreated with poly-D-lysine. Cells were equilibrated to 37 °C for 30 min before assay. OCR (pmoles/min) and ECAR (mpH/min) were measured as indicated upon cell treatment with oligomycin (0.5 mM), FCCP (0.2 mM), rotenone (1 μM), DMM (10 mM) and antimycin A (1 μM), according to the manufacturer's instructions.

Metabolome extraction. Cells were seeded at 1×10^6 cells/ml and incubated for 4 h in complete RPMI containing dialysed FBS medium. Cells were then transferred to 1.5-ml tubes and pelleted (1 min, 6,000*g*, at room temperature). Medium was removed by aspiration and the cells were washed once with 500 μl of PBS. Metabolome extraction was performed by the addition of 50 μl of ice cold solvent (40:40:20 acetonitrile:methanol:water + 0.5% formic acid). After a 5-min incubation on ice, acid was neutralized by the addition of NH_4HCO_3. After centrifugation (15 min, 16,000*g*, at 4 °C), the clean supernatant was transferred to a clean tube, frozen on dry ice and kept at -80 °C until liquid chromatography–mass spectrometry (LC–MS) analysis[20].

Succinate quantification. Wild-type CD4 T cells (1×10^6) were activated under T_H1 culture conditions. After 4 days, cells were replated into fresh medium and cultured with DMSO, 10 mM DMM, 1 mM 3NP, 100 μM TTFA or 1 μM atpenin A5 for 6 h. Cells were then collected, processed and analysed using the Succinate Assay Kit (Abcam) according to the manufacturer's protocol.

LC–MS analysis. Cell extracts were analysed using a quadrupole–orbitrap mass spectrometer (Q Exactive, Thermo Fisher Scientific) coupled to hydrophilic interaction chromatography via electrospray ionization. Liquid chromatography separation was on a XBridge BEH Amide column (2.1 mm × 150 mm, 2.5-μm particle size; Waters) using a gradient of solvent A (20 mM ammonium acetate, 20 mM ammonium hydroxide in 95:5 water:acetonitrile, pH 9.45) and solvent B (acetonitrile). Flow rate was 150 μl/min, column temperature was 25 °C, autosampler temperature was 5 °C and injection volume was 10 μl. The liquid chromatography gradient was: 0 min, 90% B; 2 min, 85% B; 3 min, 75% B; 7 min, 75% B; 8 min, 70% B; 9 min, 70% B; 10 min, 50% B; 12 min, 50% B; 13 min, 25% B; 14 min, 25% B; 16 min, 0% B; 21 min, 0% B; 22 min, 90% B; 25 min, 90% B. Autosampler temperature was 5 °C and injection volume was 10 μl. The mass spectrometer was operated in negative-ion mode to scan from *m/z* 70 to 1,000 at 1 Hz and a resolving power of 140,000[21]. Data were analysed using the MAVEN software[22].

Statistical analysis. Experiments were conducted with technical and biological replicates at an appropriate sample size, as estimated by our prior experience. No statistical methods were used to predetermine sample size. No methods of randomization and no blinding were applied. All data were replicated independently at least once as indicated in the figure legends, and all attempts to reproduce experimental data were successful. For all bar graphs, mean + s.d. are shown. All statistical analysis was performed using GraphPad Prism 7 (or more recent versions). *P* values < 0.05 were considered significant; $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$; *P* values > 0.05 were considered as non-significant. FlowJo 8.0 (or more recent versions) (Treestar) was used to analyse flow cytometry data. All sample sizes and statistical tests used are detailed in each figure legend.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The data that support the findings of this study are available from the corresponding authors upon reasonable request. RNA-seq datasets have been deposited in Gene Expression Omnibus under the accession number GSE130713.

15. Platt, R. J. et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).
16. Gagliani, N. et al. Coexpression of CD49b and LAG-3 identifies human and mouse T regulatory type 1 cells. *Nat. Med.* **19**, 739–746 (2013).
17. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
18. Anders, S., Py, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
19. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
20. Lu, W., Wang, L., Chen, L., Hui, S. & Rabinowitz, J. D. Extraction and quantitation of nicotinamide adenine dinucleotide redox cofactors. *Antioxid. Redox Signal.* **28**, 167–179 (2018).
21. Jang, C. et al. The small intestine converts dietary fructose into glucose and organic acids. *Cell Metab.* **27**, 351–361.e3 (2018).
22. Melamud, E., Vastag, L. & Rabinowitz, J. D. Metabolomic analysis and visualization engine for LC-MS data. *Anal. Chem.* **82**, 9818–9826 (2010).

Acknowledgements This work was supported by NIH grants R37 AR40072, R61AR073048 (J.C. and R.A.F.), F31 AI1333855 (J.A.S.), T32 AI7019-41 (J.A.S.), R01 CA166025-04 (L.J.M. III), T32 GM065841-14 (L.J.M. III), the Howard Hughes Medical Institute (R.A.F.), European Union's Horizon 2020, and Marie Sklodowska-Curie grant agreement no. 751423 (J.C.G.C.), and the Paradifference Foundation (L.J.M. III).

Author contributions W.B., J.A.S., J.C. and R.A.F. designed the study and wrote the manuscript. W.B. and J.A.S. designed and performed experiments. J.Z., R.Q. and Y.K. performed all bioinformatic and genomic analysis. P.B. assisted with sequencing. J.C.G.C. and J.R. designed and performed LC–MS experiments and data analysis. F.J.A.K. and L.J.M. III prepared and provided *Sdhc* cKO mouse tissue. O.K. assisted with vector cloning. H.R.S. assisted with experimentation. R.J. assisted with experimental design. All authors edited and approved the manuscript.

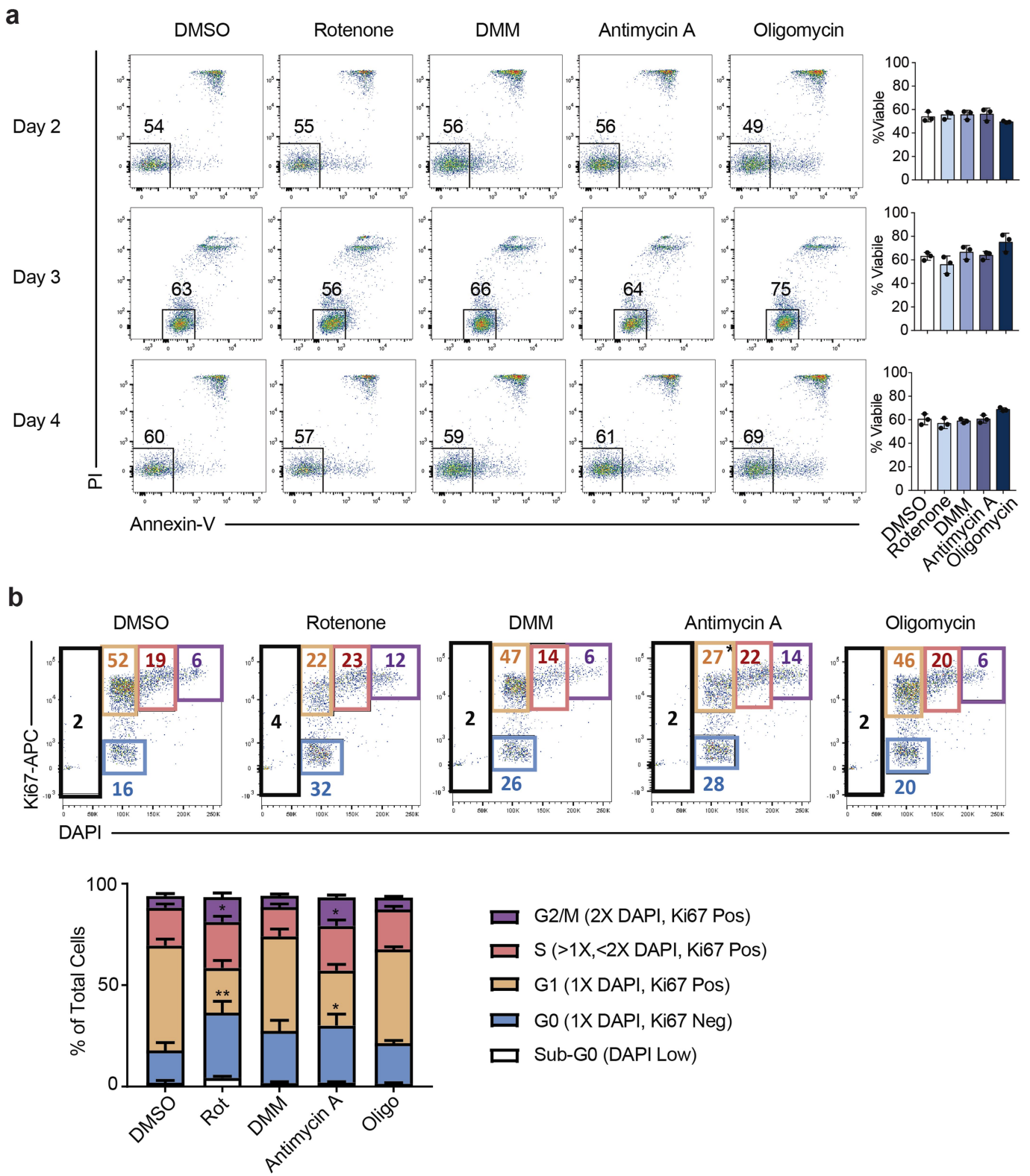
Competing interests The authors declare no competing interests.

Additional information

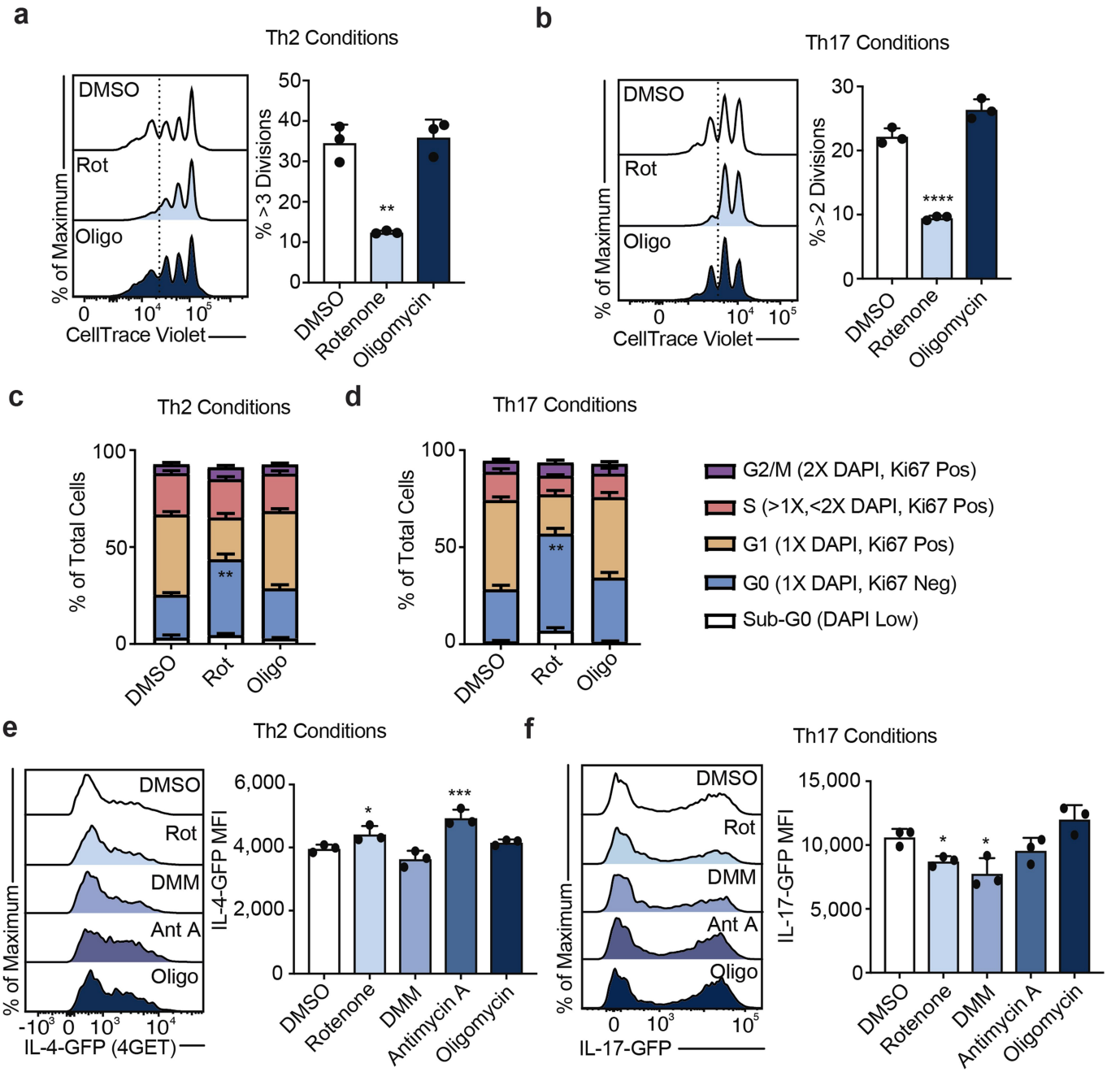
Supplementary information. is available for this paper at https://doi.org/10.1038/s41586-019-1311-3.

Correspondence and requests for materials should be addressed to J.C. or R.A.F.

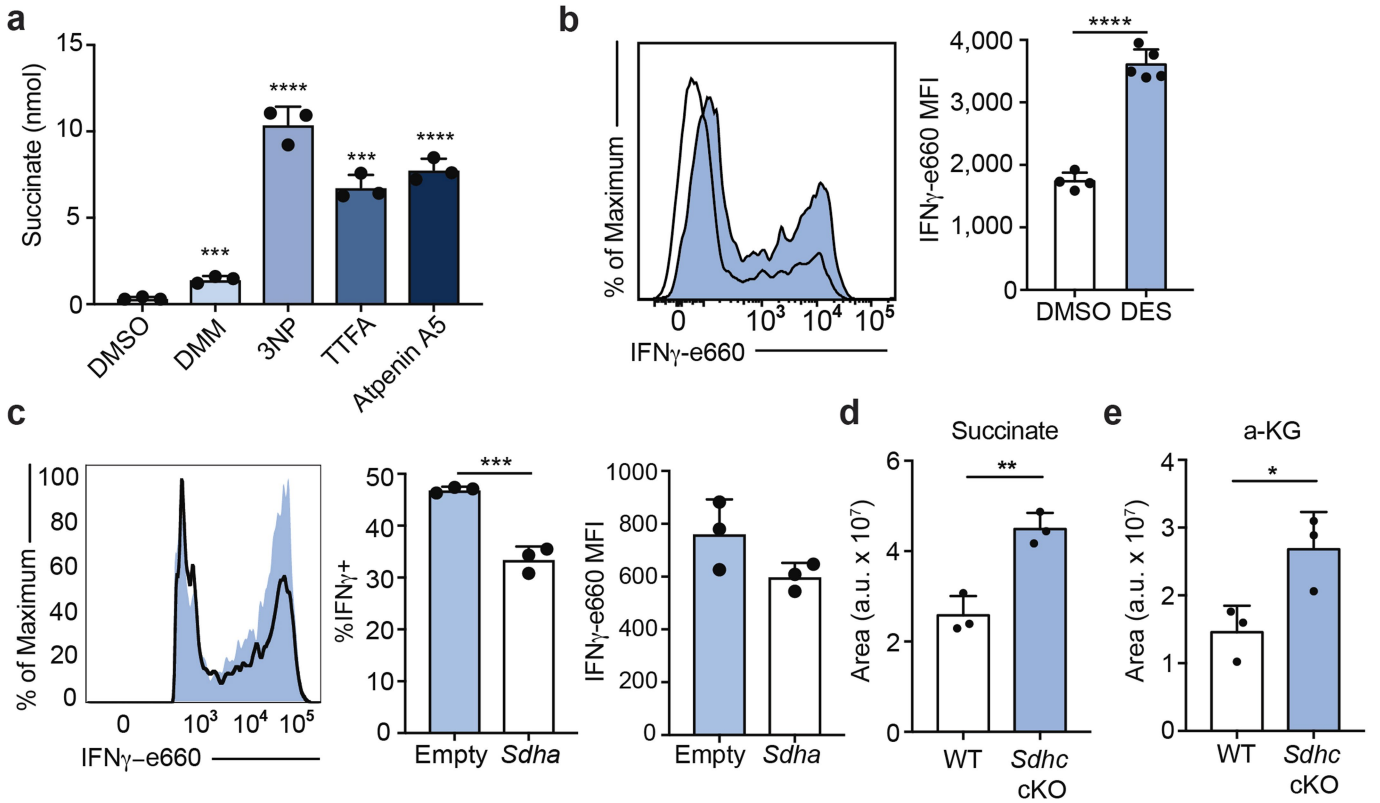
Peer review information *Nature* thanks Navdeep Chandel and the other anonymous reviewer(s) for their contribution to the peer review of this work.



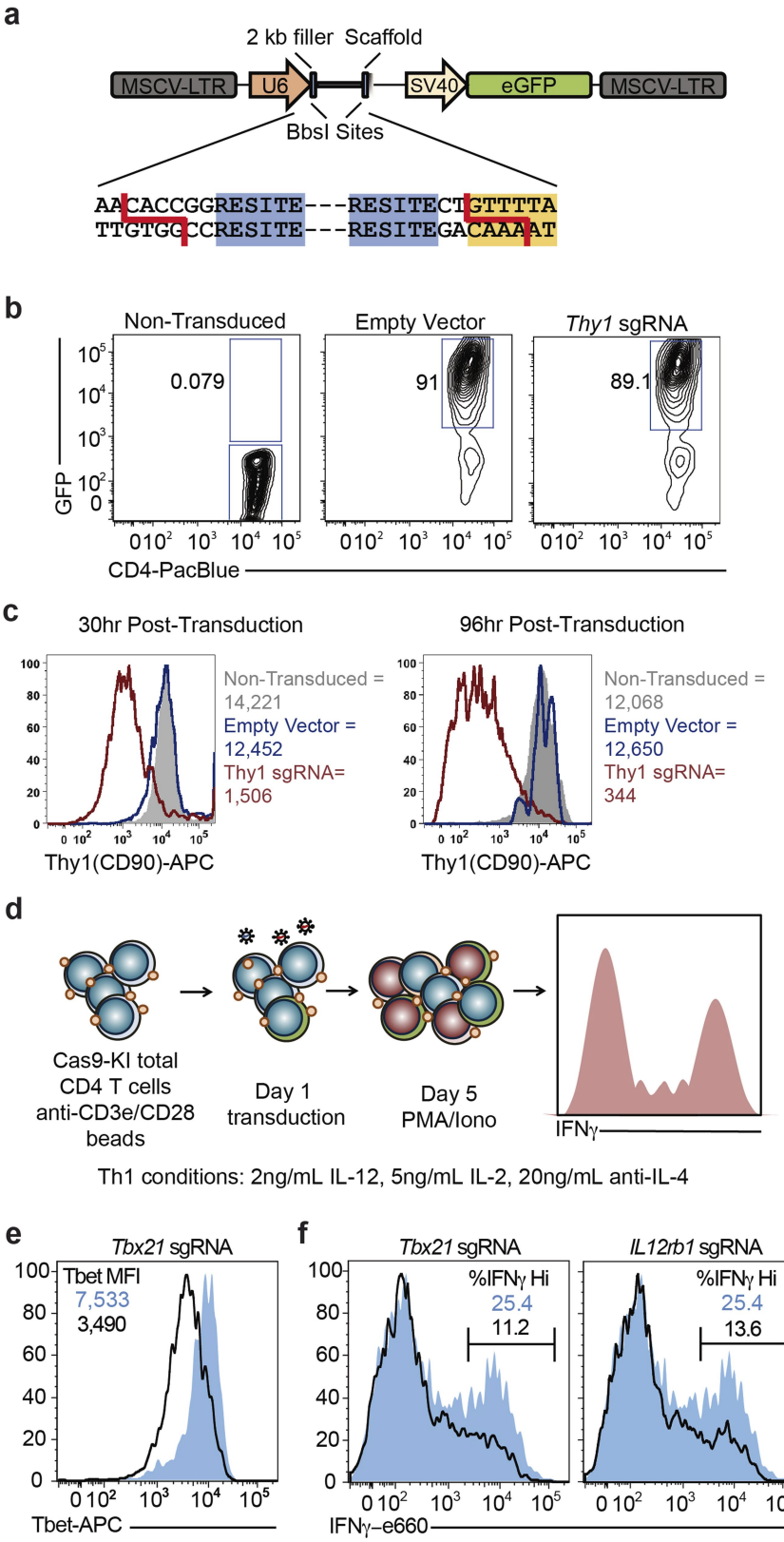
Extended Data Fig. 1 | Acute ETC inhibition alters cell cycle, but not viability, in T_H1 cells. a, Viability measured by propidium iodide (PI) and annexin-V staining of wild-type CD4 T cells cultured in T_H1 conditions and treated overnight for 16 h on day 1, 2 or 3 of culture with DMSO, rotenone, DMM, antimycin A or oligomycin ($n = 3$). **b**, Cell-cycle analysis measured by Ki67 and DAPI of CD4 T cells cultured in T_H1 conditions on day 3 following 16-h overnight treatment with DMSO ($n = 5$), rotenone, DMM, antimycin A or oligomycin ($n = 6$). *n*, number of technical replicates. Representative plots and a graph summarizing the results of three independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^*P < 0.05$, $^{**}P < 0.01$.



Extended Data Fig. 2 | ETC regulation of proliferation is conserved among subtypes of T helper cells, but ETC requirements for effector cytokine transcription differ between T_H1, T_H2 and T_H17 cells. **a**, **b**, Proliferation of wild-type CD4 T cells cultured in T_H2 (**a**) and T_H17 (**b**) conditions, following 16-h overnight treatment with DMSO, rotenone or oligomycin ($n = 3$). **c**, **d**, Cell-cycle analysis measured by Ki67 and DAPI of CD4 T cells cultured in T_H2 (**c**) and T_H17 (**d**) conditions on day 3 following 16-h overnight treatment with DMSO, rotenone, DMM, antimycin A or oligomycin ($n = 6$). **e**, **f**, Effector cytokine transcription after PMA and ionomycin restimulation at day 5 measured by IL-4–GFP (4GET) reporter expression in cells cultured in T_H2 conditions (**e**) and IL17–GFP reporter expression in cells cultured in T_H17 conditions (**f**) following 16-h overnight treatment with DMSO, rotenone, DMM, antimycin A or oligomycin ($n = 3$). *n*, number of technical replicates. Representative plots and a graph summarizing the results of three independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

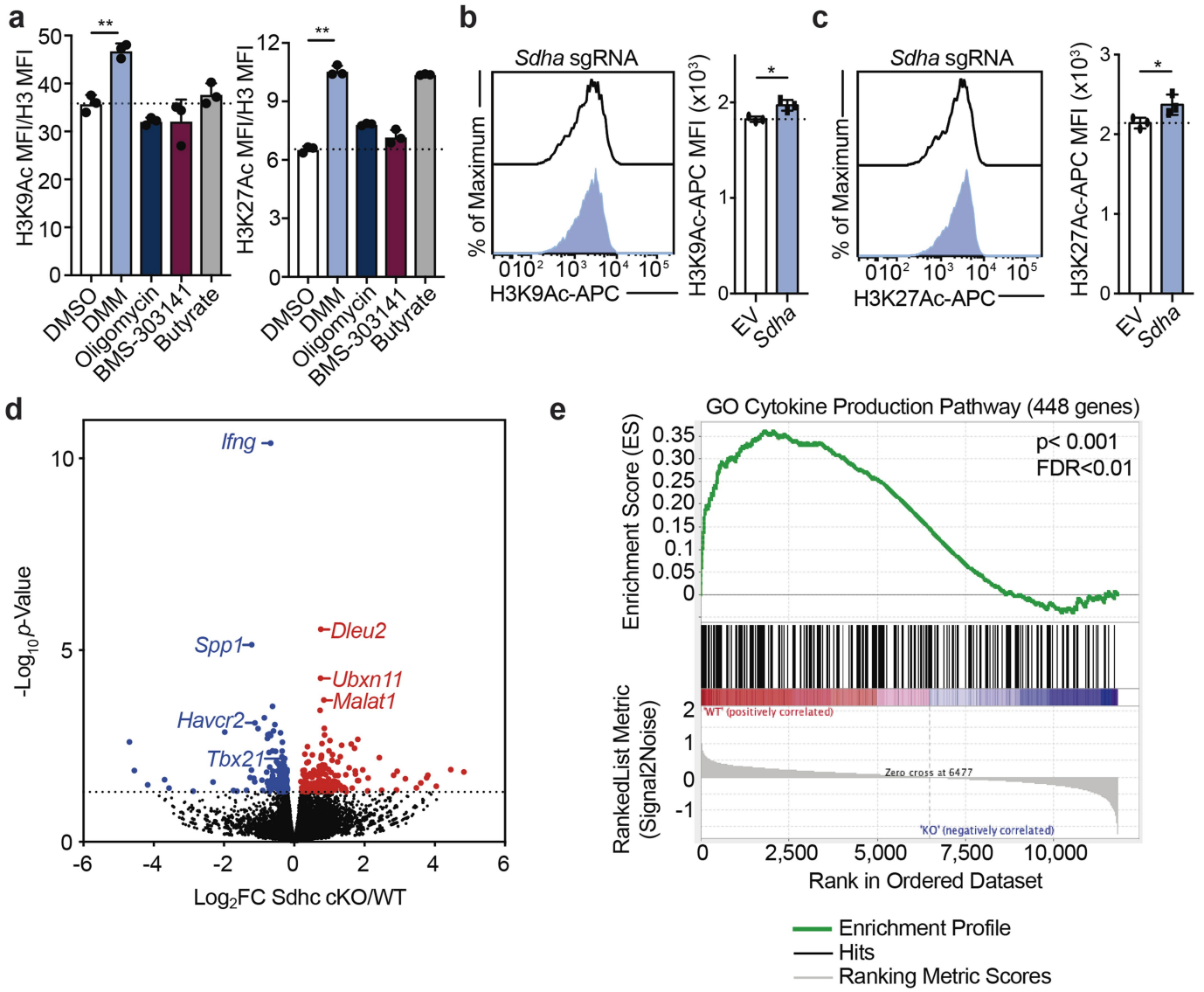


Extended Data Fig. 3 | Complex II inhibition is functional and leads to a loss of IFNγ production in T_H1 cells. a, Cellular succinate at day 5 evaluated using Succinate Assay Kit (Abcam) in wild-type CD4 T cells cultured in T_H1 conditions following 6-h treatment with DMSO, 10 mM DMM, 1 mM 3NP, 100 μM TTFA or 1 μM atpenin A5 ($n = 3$). **b**, IFNγ protein production after PMA and ionomycin restimulation at day 5 of wild-type CD4 T cells cultured in T_H1 conditions following 16-h overnight treatment with 10 mM diethyl succinate (DES) ($n = 5$) or DMSO ($n = 4$). **c**, IFNγ protein production after PMA and ionomycin restimulation at day 5 of Cas9-expressing CD4 T cells cultured in T_H1 conditions transduced with one of three individual sgRNA targeting *Sdha*, or an empty-vector control ($n = 3$ biological replicates). **d**, **e**, Total cellular succinate (**d**) and α-ketoglutarate (**e**) measured by LC–MS analysis in wild-type or *Sdhc* cKO CD4 T cells cultured in T_H1 conditions after 4-h culture in dialysed FBS-containing medium at day 5 ($n = 3$). *n*, number of technical replicates, unless otherwise stated. Representative plots and a graph summarizing the results of at least two independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^*P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, $^{****}P < 0.0001$.

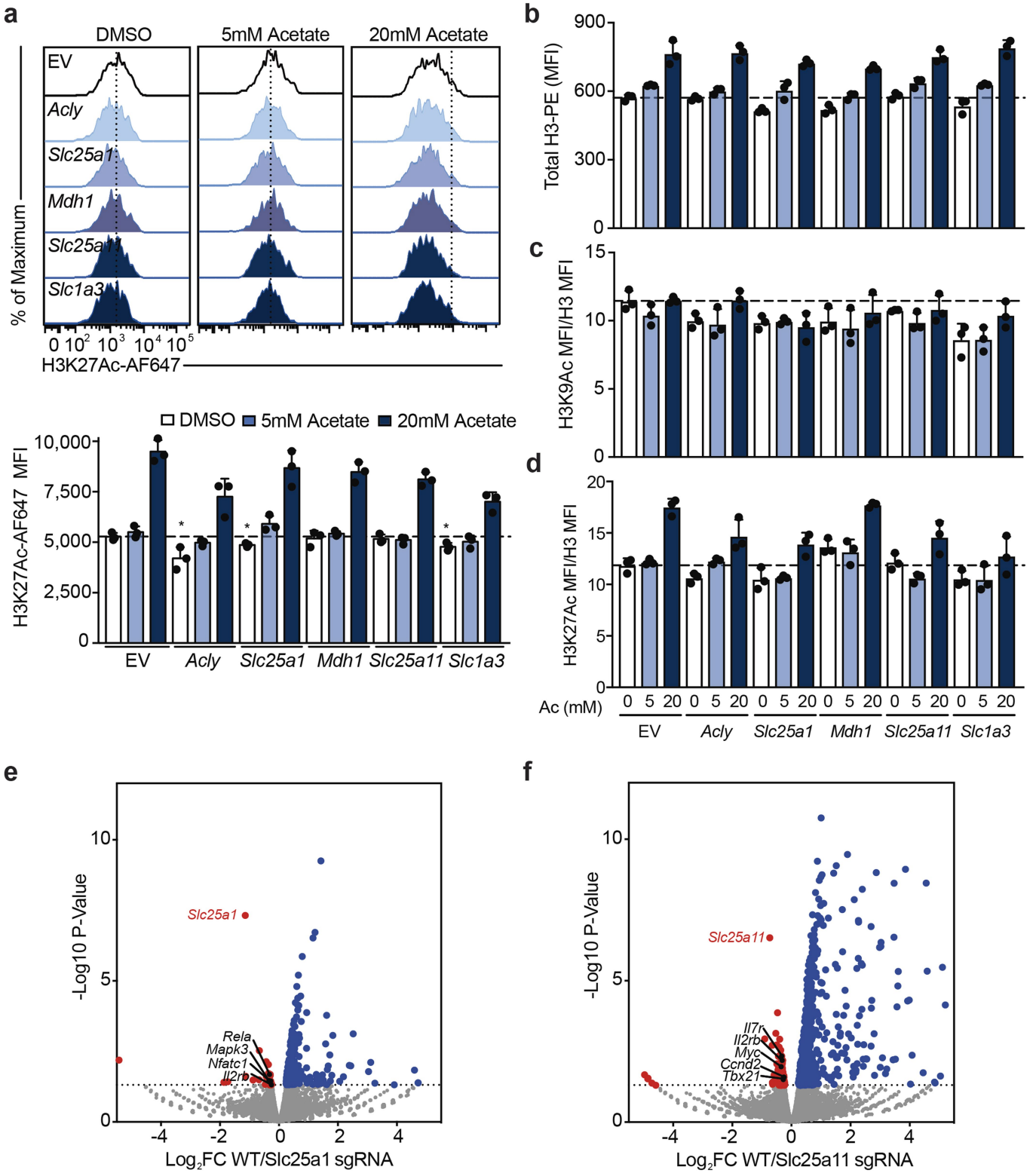


Extended Data Fig. 4 | See next page for caption.

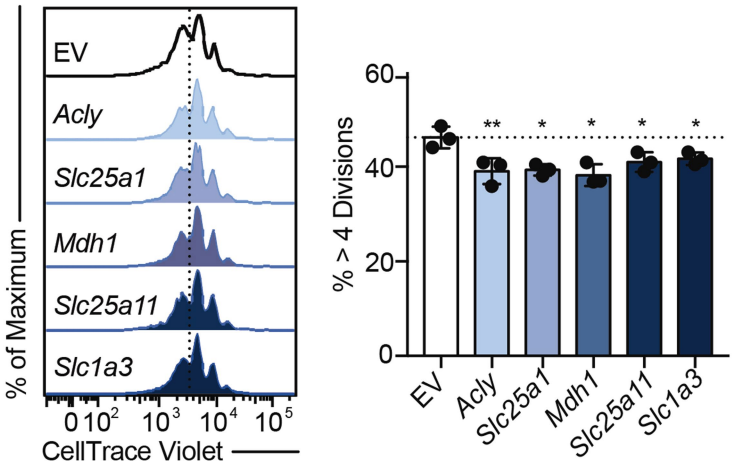
Extended Data Fig. 4 | Retroviral expression of sgRNA in Cas9-expressing CD4 T cells. a, Schematic of MG-Guide retroviral vector. **b**, CD4 T cells from Cas9-expressing mice were stimulated with anti-CD3 and anti-CD28 coated beads for 24 h and retrovirally transduced with either a MG-Guide (empty vector) or a MG-Guide vector cloned to express a sgRNA against *Thy1* (*Thy1* sgRNA). GFP expression was measured at 24 h after transduction, compared to non-transduced cells. **c**, THY1.1 protein expression was measured in transduced (empty vector blue line; *Thy1* sgRNA red line) and non-transduced (solid grey) cells by flow cytometry at 30 and 96 h after transduction. **d**, Schematic of experimental design for functional T_H1 sgRNA studies. **e**, CD4 T cells from Cas9-expressing mice were stimulated with anti-CD3 and anti-CD28 beads in IL-2 (5 ng ml^{-1}), anti-IL-4 (10 μg ml^{-1}) and IL-12 (2 ng ml^{-1}) and retrovirally transduced 24 h after activation with either empty MG-Guide (shaded blue) or MG-Guide expressing a sgRNA against *Tbx21* (outline). TBET protein expression was measured by intracellular flow cytometry on day 3. **f**, Cas9-expressing CD4 T cells were cultured as above, and infected with MG-Guide, a sgRNA against *Tbx21* or a sgRNA against *Il12rb*. IFNγ protein was measured by intracellular flow cytometry on day 5 after restimulation with PMA (20 ng ml^{-1}) and ionomycin (1 μg ml^{-1}).



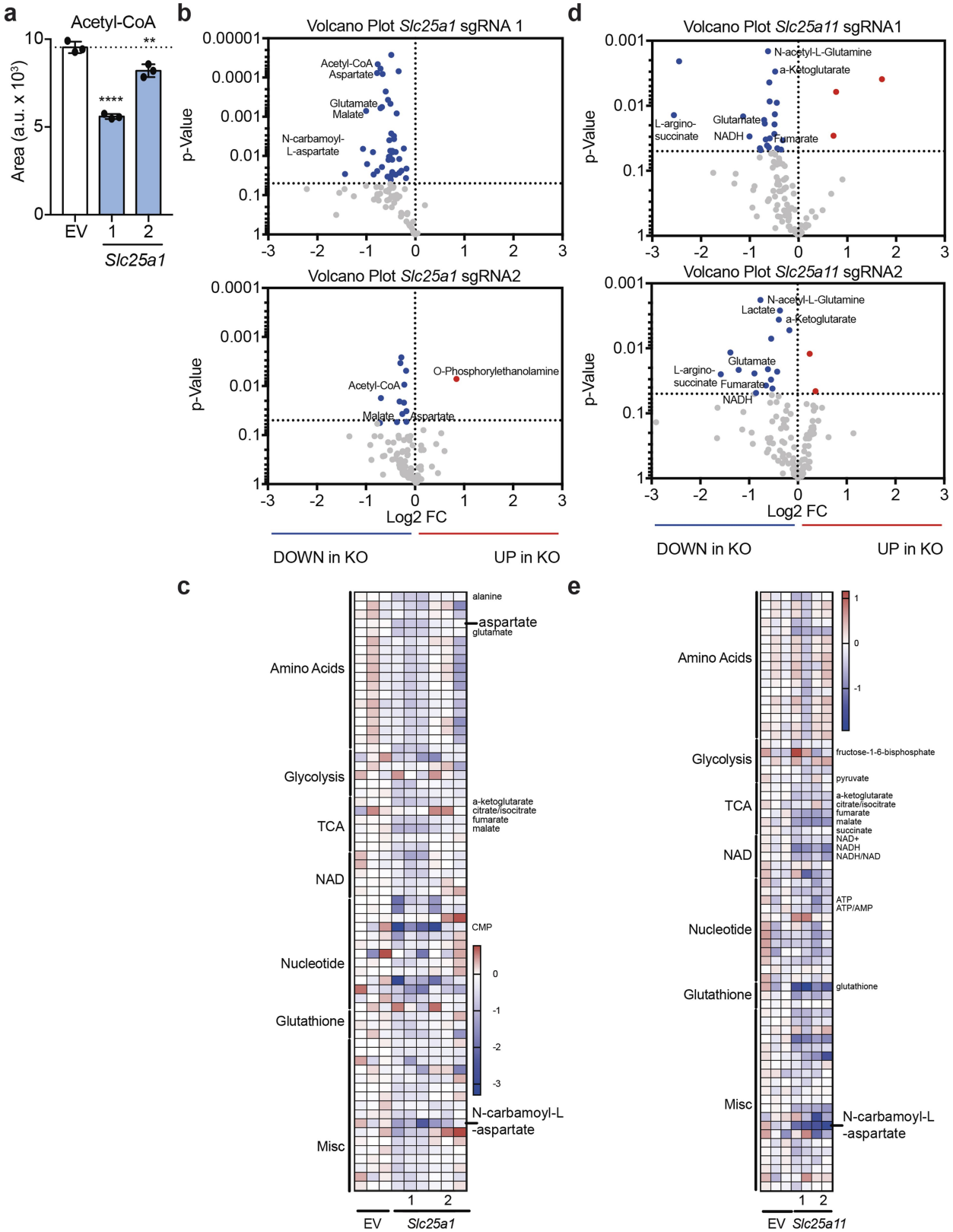
Extended Data Fig. 5 | Complex II regulates epigenetic modifications and program-specific gene expression in T_H1 cells. a, H3K9 acetylation and H3K27 acetylation, normalized to total cellular H3 and 1× DNA content on day 3 of wild-type CD4 T cells cultured in T_H1 conditions after 16-h overnight treatment with DMSO, DMM, oligomycin, BMS-303141 or butyrate ($n = 3$). **b**, **c**, H3K9 acetylation (**b**) and H3K27 acetylation (**c**) at day 5 of Cas9-expressing CD4 T cells cultured in T_H1 conditions transduced with one of three individual sgRNA targeting *Sdha*, or an empty-vector control ($n = 3$ biological replicates). **d**, **e**, Volcano plot summarizing RNA-seq data that indicate the most-differentially regulated transcripts between wild-type and *Sdhc* cKO T_H1 cells (**d**) and gene set enrichment analysis (GSEA) plot of the GO cytokine production pathway (**e**) ($n = 3$ biological replicates). *n*, number of technical replicates unless otherwise stated. Representative plots and a graph summarizing the results of at least two independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^*P < 0.05$, $^{**}P < 0.01$.



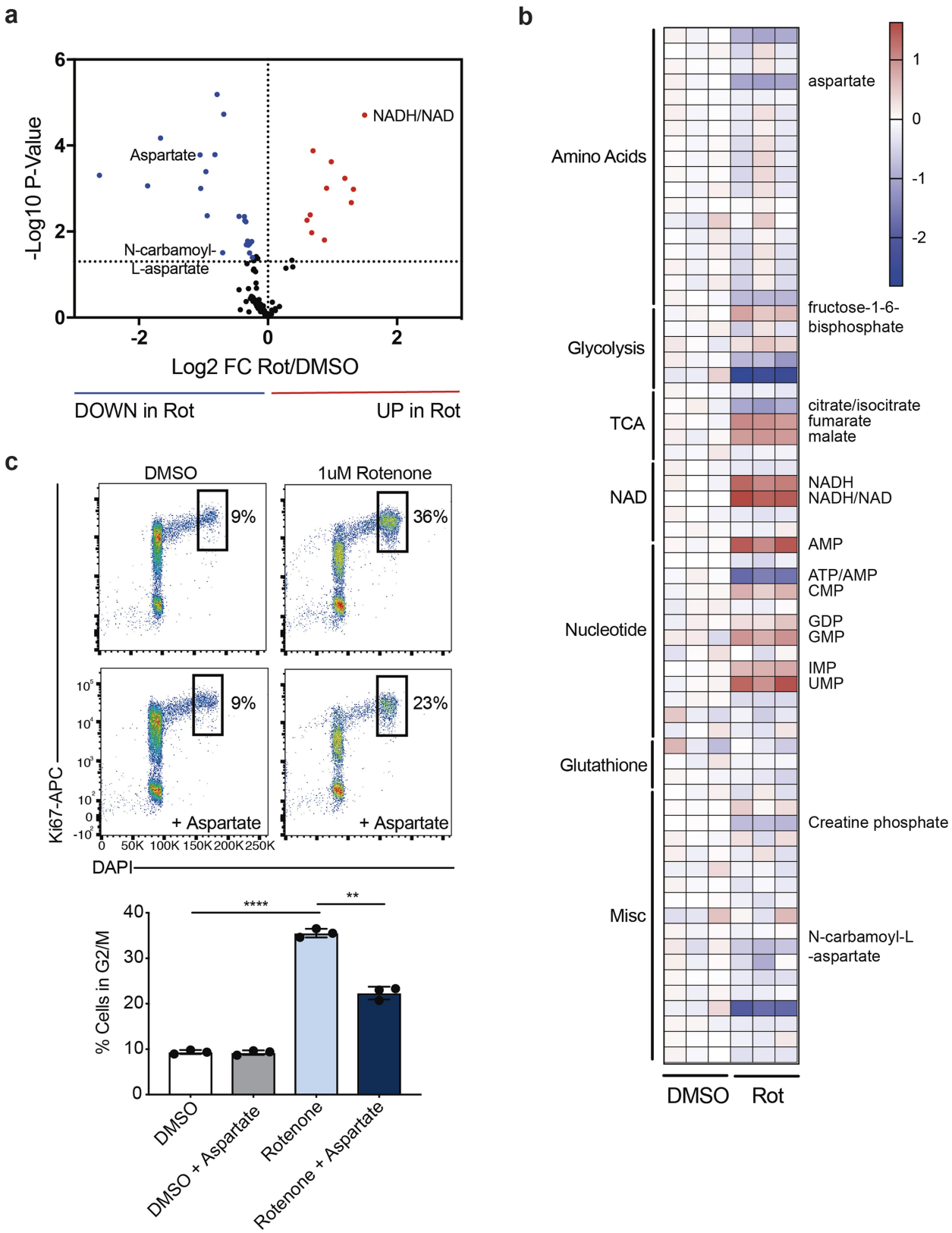
Extended Data Fig. 6 | The malate–aspartate shuttle and mitochondrial citrate export dynamically regulate histone acetylation and program-specific gene expression in T_H1 cells. a–d, H3K27 acetylation (**a**), total cellular H3 (**b**), H3K9 acetylation normalized to total cellular H3 and 1× DNA content (**c**) and H3K27 acetylation normalized to total cellular H3 and 1× DNA content (**d**) on day 4 of Cas9-expressing CD4 T cells transduced with three individual sgRNAs targeting *Acly*, *Slc25a1*, *Mdh1*, *Slc25a11* or *Slc1a3*, or empty vector, cultured in T_H1 conditions ($n = 3$ biological replicates). **e**, **f**, Volcano plot summarizing RNA-seq data that indicate the most-differentially regulated transcripts at day 5 in Cas9-expressing CD4 T cells cultured in T_H1 conditions transduced with empty vector or one sgRNA targeting *Slc25a1* (**e**) or *Slc25a11* (**f**) ($n = 2$ biological replicates). Representative plots and a graph summarizing the results of at least two independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^*P < 0.05$.



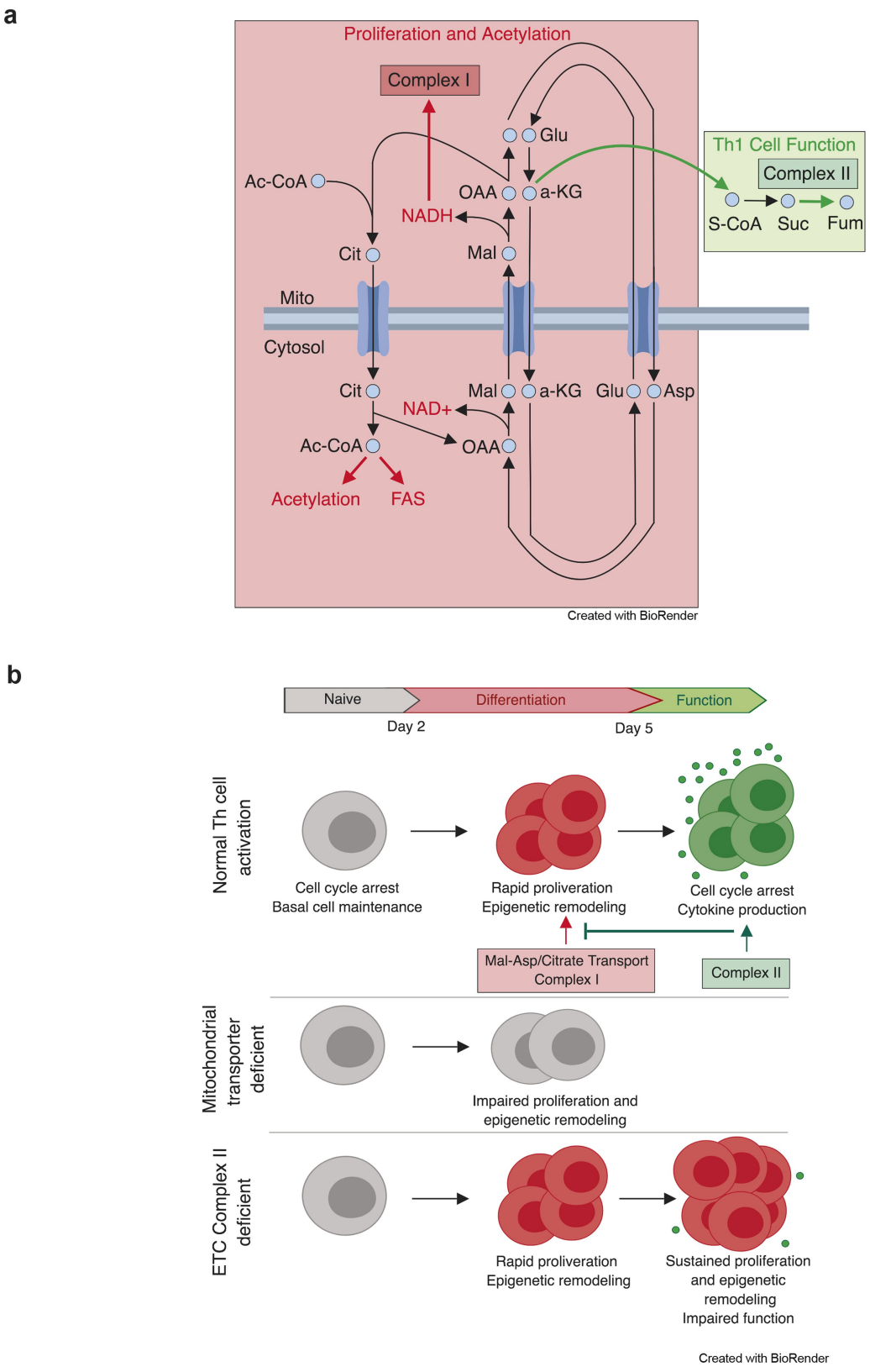
Extended Data Fig. 7 | The malate–aspartate shuttle and mitochondrial citrate export are required for proliferation in T_H1 cells. Proliferation of Cas9-expressing CD4 T cells transduced with empty-vector control or one of three individual sgRNAs targeting *Acly*, *Slc25a1*, *Mdh1*, *Slc25a11* or *Slc1a3*, cultured in T_H1 conditions at day 5 ($n = 3$ biological replicates). Representative plots and a graph summarizing the results of at least two independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^*P < 0.05$, $^{**}P < 0.01$.



Extended Data Fig. 8 | The malate–aspartate shuttle and mitochondrial citrate export regulate levels of cellular acetyl-CoA and cellular metabolism. a, Cellular acetyl-CoA measured by LC–MS analysis in Cas9-expressing CD4 T cells transduced with empty vector or two individual sgRNAs targeting *Slc25a1*, as described, on day 5 of culture in T_H1 conditions ($n = 2$ biological replicates, $n = 3$ technical replicates). a.u., arbitrary units. **b**, **c**, Volcano plot (**b**) and heat map (**c**) of all metabolites measured by LC–MS analysis in Cas9-expressing CD4 T cells transduced with empty vector or two individual sgRNAs targeting *Slc25a1*, as described, on day 5 of culture in T_H1 conditions ($n = 2$ biological replicates, $n = 3$ technical replicates). **d**, **e**, Volcano plot (**d**) and heat map (**e**) of all metabolites measured by LC–MS analysis in Cas9-expressing CD4 T cells transduced with empty vector or two individual sgRNAs targeting *Slc25a1*, as described, on day 5 of culture in T_H1 conditions ($n = 2$ biological replicates, $n = 2$ technical replicates). Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^{**}P < 0.01$, $^{****}P < 0.0001$.



Extended Data Fig. 9 | Complex I activity is required for aspartate production and cell-cycle progression in activating T_H1 cells. **a**, **b**, Volcano plot (**a**) and heat map (**b**) of all metabolites measured by LC–MS analysis in wild-type CD4 T cells treated acutely for 4 h on day 5 of culture in T_H1 conditions ($n = 3$). **c**, Cell-cycle analysis using Ki67 and DAPI of wild-type CD4 T cells cultured in T_H1 conditions at day 3 following 16-h overnight treatment with DMSO or rotenone ± 20 mM aspartate ($n = 3$). *n*, number of technical replicates. Representative plots and a graph summarizing the results of at least two independent experiments are shown. Mean and s.d. of replicates are presented on summarized plots and unpaired, two-tailed *t*-test used to determine significance. $^{**}P < 0.01$, $^{****}P < 0.0001$.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Conceptual models of mitochondrial metabolite transport and the consequence of metabolic perturbations on T_H1 cell activation. a, Early-stage T_H1 cell activation is supported by the malate–aspartate shuttle and mitochondrial citrate export. These mitochondrial transport systems provide the key substrates that are needed for cell division and histone acetylation. Citrate export results in the production of cytosolic acetyl-CoA that can be used to synthesize the fatty acids that are needed for plasma membrane expansion during division, as well as the acetyl groups that are used for histone acetylation. Interconnected with this export pathway is the malate–aspartate shuttle, a carbon-neutral cycle that results in the net movement of NAD^+ to the cytosol and NADH into the mitochondria, through which the cycle can fuel the activity of ETC complex I. Through the activity of complex I, NAD^+ can be continually recycled, which enables the production of aspartate (an essential precursor for nucleotide synthesis). These processes are antagonized by the activity of SDH (ETC complex II), which consumes α-ketoglutarate; this limits the availability of the latter for the malate–aspartate shuttle and promotes effector functions of T_H1 cells. **b**, T-helper-cell activation is defined by two major phases: (1) a period of rapid division and epigenetic remodelling, and (2) cell-cycle arrest and cytokine production. Each of these phases is supported by a discrete component of mitochondrial metabolism. The malate–aspartate shuttle and mitochondrial citrate export generate the material needed for early-phase cell differentiation to occur. As differentiation continues, the activity of complex II draws carbon away from the shuttle, and thus acts to pull activated T_H1 cells out of the differentiation process and to enable them to fully engage their terminal effector cell program. When the mitochondrial transport networks are disrupted, T_H1 cells are unable to properly proliferate or epigenetically reprogram. By contrast, inhibiting the activity of complex II causes activated T_H1 cells to continuously proliferate and remodel their chromatin, which prevents them from exiting the differentiation phase and engaging their terminal effector program.

nature research

Corresponding author(s): Flavell RA, Craft J

Last updated by author(s): May 15, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	*Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.*
Data analysis	For flow cytometry software information, please see the flow cytometry form. All graphing and associated statistical analysis was performed using GraphPad Prism 7 or newer. MAVEN (build 682) was used for LC-MS analysis (http://genomics-pubs.princeton.edu/mzroll/index.php). STAR 2.7.0 was used for alignment of RNA-seq data. HTseq 0.11.1 was used for gene expression. DEseq2 was used for differential expression.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request. RNA-seq data sets have been deposited in Gene Expression Omnibus under the accession number GSE130713

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size for experimentation. Given the minimal experimental variation in T cell assays using inbred mouse strains as a cell source, a minimum of three technical replicates were used per sample.
Data exclusions	sgRNA transduction experiments were evaluated based on the efficacy of positive and negative controls. Experiments for which the positive or negative controls did not yield expected results were excluded. Predetermined exclusion criteria indicated that any experiment for which transduction with Tbx21 targeting sgRNA (experimental positive control) did not yield a significant ($p<0.05$) decrease in IFNg expression was to be excluded. No other data were excluded.
Replication	Using the data quality parameters described above, all experiments shown in the manuscript were reliably reproducible
Randomization	Because all cells used for analysis were from the same inbred strain of mouse, no randomization was performed.
Blinding	No blinding was performed. Blinding was unnecessary as all data collection and analysis is quantitative and not qualitative in nature.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	anti-CD4 Pe-Cy7 Biolegend Cat#100422 Clone#GK1.5 Lot#B224943 (1:1000) anti-Tbet e660 eBioscience Cat#50-5825-80 Clone#eBio4b10 Lot#e12135-1634 (1:50) anti-CD90.2 APC BD Cat#553007 Clone#53-2.1 Lot#72965 (1:500) Anti-Ifng BV421 Biolegend Cat#505829 Clone#XMG1.2 Lot#B214227 (1:400) Anti-Ifng e660 eBioscience Ref#50-7311-82 Clone#XMG1.2 Lot#e15675-103 (1:400) anti-GFP FITC Life Technologies Cat#A21311 Polyclonal Lot#1567217 (1:200) anti-mouse Ki-67 Biolegend Cat#652422 Clone 16A8 (1:200) anti-mouse H3 Cell Signaling Cat#14269S Clone 1B1B2 (1:200) anti-mouse Acetyl-histone H3(lys9) Cat#9649S Clone C5B11 (1:50) anti-mouse Acetyl-histone H3 (Lys27) Cat#8173S Clone D5E4 (1:50)
Validation	All antibodies listed have been used by a large number of laboratories including our own for many publications, including those in Nature. Each antibody is validated by the company of purchase, and dilution is determined experimentally for each clone individually.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	293T (ATCC CRL-3216)
Authentication	Cell lines obtained directly from ATCC. No cell line authentication performed.
Mycoplasma contamination	Cell lines were not tested for mycoplasma
Commonly misidentified lines (See ICLAC register)	The cell line used is not listed in the ICLAC database.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mus musculus, B6, sex-matched male and female, 6-8 weeks old
Wild animals	n/a
Field-collected samples	n/a
Ethics oversight	All mice required for this study were housed and maintained under specific-pathogen-free conditions in the animal facility of the Yale University School of Medicine, and all corresponding animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) of Yale University. This study was conducted in compliance with all relevant ethical regulations.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were harvested, washed at least once with PBS, and stained with antibodies targeting surface antigens in 2%FBS 1mM EDTA in PBS for 30 minutes on ice. For analysis of intracellular cytokine staining, cells were then fixed and permeabilized with fixation/ permeabilization solution (BD Cytofix/Cytoperm Cat#554714) at 4C for 20 minutes. For nuclear transcription factor staining, cells were fixed and permeabilized with eBioscience Foxp3/Transcription Factor Fixation/Permeabilization Concentrate and Diluent (Cat #00-5521-00). Cells were then washed three times with 1X permeabilization buffer (eBioscience Permeabilization Buffer Cat #00-8333-56) and stained for intracellular antigens for 1 hour on ice in 1X permeabilization buffer. Cells were then washed two times before proceeding to cytometry analysis. For cells not transduced with retrovirus, cells were labeled with LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (ThermoFisher Cat# L34965) per the manufactures protocol before surface staining. Further details regarding specific staining protocols and antibodies used can be found in the methods section of the manuscript. Cells were harvested, washed at least once with PBS, and stained with antibodies targeting surface antigens in 2%FBS 1mM EDTA in PBS for 30 minutes on ice. For analysis of intracellular cytokine staining, cells were then fixed and permeabilized with fixation/ permeabilization solution (BD Cytofix/Cytoperm Cat#554714) at 4C for 20 minutes. For nuclear transcription factor staining, cells were fixed and permeabilized with eBioscience Foxp3/Transcription Factor Fixation/Permeabilization Concentrate and Diluent (Cat #00-5521-00). Cells were then washed three times with 1X permeabilization buffer (eBioscience Permeabilization Buffer Cat #00-8333-56) and stained for intracellular antigens for 1 hour on ice in 1X permeabilization buffer. Cells were then washed two times before proceeding to cytometry analysis. For cells not transduced with retrovirus, cells were labeled with LIVE/DEAD Fixable Aqua Dead Cell Stain Kit (ThermoFisher Cat# L34965) per the manufactures protocol before surface staining. Further details regarding specific staining protocols and antibodies used can be found in the methods section of the manuscript.
Instrument	BD LSRII custom order product
Software	BD FACSDiva Software was used to collect raw data files from all flow cytometry experiments. All resultant data files were analyzed using FlowJo version 8 or newer.
Cell population abundance	Cells were purified for three replicate screens using a BD FACS Aria as described in methods. Due to limited cell numbers, all resultant cells were processed for gDNA isolation and sequencing.

Purity of magnetic bead CD4+ T cell isolation was determined by flow cytometry with >90% CD4 expressing cells used for all experiments.

Gating strategy

For all experiments measuring Ifng-Kat reporter expression, cells were gated on FSC-A high, DAPI low >> lymphocyte FSC-A, SSC-a >> CD4 positive, GFP positive based on clear distinctions between populations in the pseudo-color plot. Cells were binned into 4 quartiles of katushka expression. This gating strategy is shown in Extended Data Figure 4a.

For experiments measuring intracellular cytokine in retrovirally transduced cells, gating was as followed: FSC-A, SSC-A live cells >> SSC-W, SSC-H singlets >> FSC-H, FSC-W singlets >> CD4 positive, GFP positive live transduced cells.

For experiments measuring intracellular cytokine in cells not transduced, gating was as followed: FSC-A, SSC-A live cells >> SSC-W, SSC-H singlets >> FSC-H, FSC-W singlets >> CD4 positive, LIVE/DEAD Negative.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.

LETTER

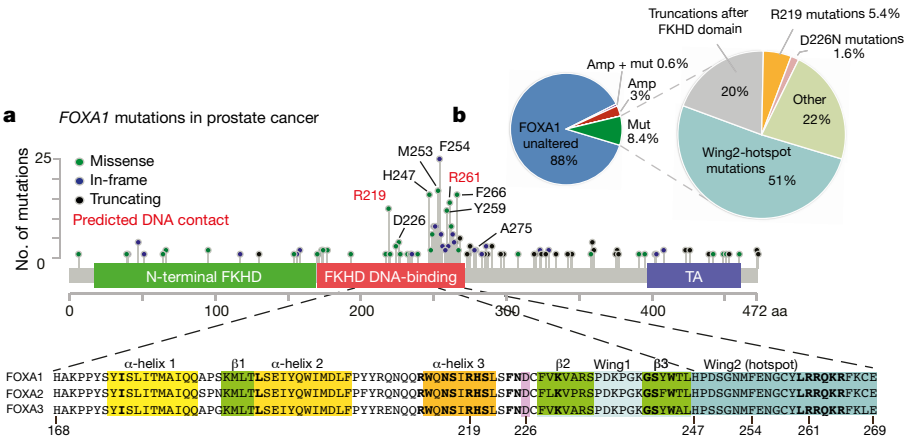
https://doi.org/10.1038/s41586-019-1318-9

FOXA1 mutations alter pioneering activity, differentiation and prostate cancer phenotypes

Elizabeth J. Adams[1], Wouter R. Karthaus[1], Elizabeth Hoover[1], Deli Liu[2,3,4], Antoine Gruet[5], Zeda Zhang[1,6], Hyunwoo Cho[7,8], Rose DiLoreto[8,9], Sagar Chhangawala[7,8], Yang Liu[10], Philip A. Watson[1], Elai Davicioni[10], Andrea Sboner[2,4,11], Christopher E. Barbieri[2,3,11], Rohit Bose[12], Christina S. Leslie[8] & Charles L. Sawyers[1,13]*

Mutations in the transcription factor FOXA1 define a unique subset of prostate cancers but the functional consequences of these mutations and whether they confer gain or loss of function is unknown[1–9]. Here, by annotating the landscape of *FOXA1* mutations from 3,086 human prostate cancers, we define two hotspots in the forkhead domain: Wing2 (around 50% of all mutations) and the highly conserved DNA-contact residue R219 (around 5% of all mutations). Wing2 mutations are detected in adenocarcinomas at all stages, whereas R219 mutations are enriched in metastatic tumours with neuroendocrine histology. Interrogation of the biological properties of wild-type FOXA1 and fourteen *FOXA1* mutants reveals gain of function in mouse prostate organoid proliferation assays. Twelve of these mutants, as well as wild-type FOXA1, promoted an exaggerated pro-luminal differentiation program, whereas two different R219 mutants blocked luminal differentiation and activated a mesenchymal and neuroendocrine transcriptional program. Assay for transposase-accessible chromatin using sequencing (ATAC-seq) of wild-type FOXA1 and representative Wing2 and R219 mutants revealed marked, mutant-specific changes in open chromatin at thousands of genomic loci and exposed sites of FOXA1 binding and associated increases in gene expression. Of note, ATAC-seq peaks in cells expressing R219 mutants lacked the canonical core FOXA1-binding motifs (GTAAAC/T) but were enriched for a related, non-canonical motif (GTAAAG/A), which was preferentially activated by R219-mutant FOXA1 in reporter assays. Thus, *FOXA1* mutations alter its pioneering function and perturb normal luminal epithelial differentiation programs, providing further support for the role of lineage plasticity in cancer progression.

To investigate the role of mutant and wild-type *FOXA1* in prostate cancer, we examined the landscape of *FOXA1* mutations across a cohort of 3,086 patients with primary or metastatic disease. The overall frequency of *FOXA1* mutation in these patients is around 11% (Fig. 1a, b), 3% of which are genomic amplifications and 8.4% are somatic point mutations, with less than 1% having both types of mutations (Fig. 1b). More than 50% of *FOXA1* mutations map to a specific hotspot in the Wing2 region of the forkhead (FKHD) DNA-binding domain, often as missense mutations or indels in Wing2 (mainly between H247 and



c

MSK-IMPACT 504 mutant frequencies by stage	R219	Wing2-hotspot alterations	Others	Total
Locoregional (*n* = 30)	6.7% (2)	66.7% (20)	26.7% (8)	100%
Metastatic non-castration-resistant (*n* = 13)	0.0% (0)	38.5% (5)	61.5% (8)	100%
Castration-resistant (*n* = 18)	11.1% (2)	44.4% (8)	44.4% (8)	100%

d

Point mutation frequencies in adenocarcinoma vs NEPC	*FOXA1* mutant cases			Non-*FOXA1* mutant cases
	R219 point mutants	Other point mutants	Total	
Adenocarcinoma (*n* = 1,742)	9.5% (8)	90.5% (76)	100% (84)	1,658
NEPC (*n* = 80)	75% (3)	25% (1)	100% (4)	76

***

Fig. 1 | Recurrent *FOXA1* mutations in prostate cancer cluster in the FKHD DNA-binding domain. a, Top, Distribution of *FOXA1* mutations from a pan-prostate cancer analysis of 3,086 patients along the protein sequence, depicting the various alterations seen in patients. Bottom, The amino acid sequence of the conserved FKHD DNA-binding domain, with secondary structural elements indicated. Residues in red (top) or bold (bottom) are predicted to make contacts with DNA[10]. **b**, Classification of observed *FOXA1* alterations. Mutations can be subdivided into several classes on the basis of their location in the FOXA1 protein. Amp, amplification; mut, somatic point mutation. **c**, Frequency of the various classes of *FOXA1* alterations in the three clinical stages reported in MSK-IMPACT 504. Data are expressed as percentage of the total number of samples with *FOXA1* mutations at a given clinical stage. **d**, Prevalence of R219 mutations compared to all other point mutations found in *FOXA1* in adenocarcinoma versus NEPC. Cases pooled from the Trento–Cornell–Broad[12] dataset and MSK-IMPACT 1708. ***$P = 0.0059$, Fisher's exact test, two-sided.

[1]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [2]Sandra and Edward Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA. [3]Department of Urology, Weill Cornell Medicine, New York, NY, USA. [4]HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine, Weill Cornell Medical College, New York, NY, USA. [5]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [6]Louis V. Gerstner Jr Graduate School of Biomedical Sciences, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [7]Computational and Systems Biology Program, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [8]Physiology, Biophysics, and Systems Biology Program, Weill Cornell Graduate School, New York, NY, USA. [9]Tri-Institutional Training Program in Computational Biology & Medicine, Weill Cornell Medicine, New York, NY, USA. [10]GenomeDx Bioscience, Vancouver, British Columbia, Canada. [11]Englander Institute for Precision Medicine of Weill Cornell Medicine and NewYork-Presbyterian Hospital, New York, NY, USA. [12]Departments of Anatomy, Medicine and Urology, University of California, San Francisco, San Francisco, CA, USA. [13]Howard Hughes Medical Institute, Chevy Chase, MD, USA. *e-mail: sawyersc@mskcc.org

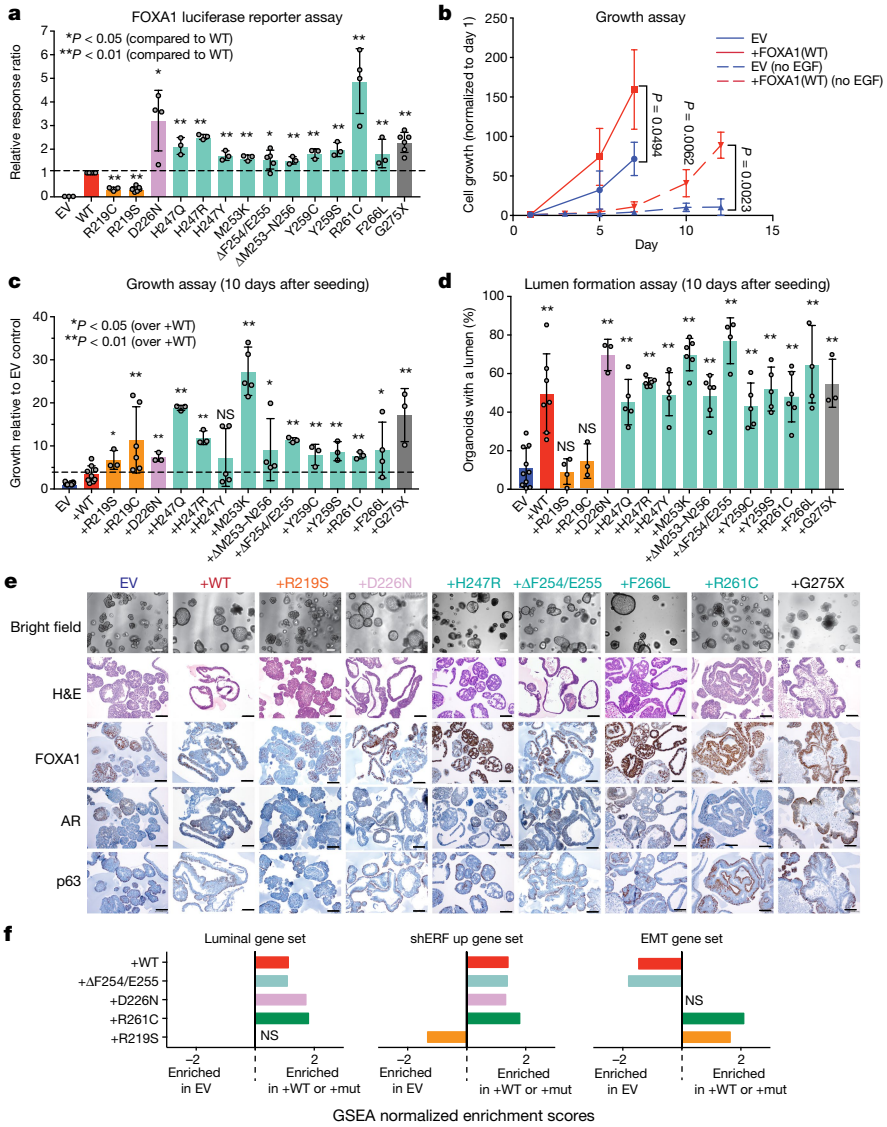


Fig. 2 | Expression of *FOXA1* mutants promotes growth and reveals distinct morphologies for different classes of alterations. a, FOXA1 luciferase reporter assay with results normalized to level of FOXA1(WT) activity. Colours indicate the position of the altered amino acid within the FKHD DNA-binding domain depicted in Fig. 1a. Grey indicates truncation. **b**, Overexpression of FOXA1 promotes growth in prostate organoids in standard medium conditions (solid lines, $n = 3$ biological replicates) and in restrictive medium conditions (dashed lines, no EGF, $n = 8$ biological replicates). EV, pCW empty vector control. **c**, Overexpression of wild-type (+WT) or various *FOXA1* mutants promotes growth ten days after seeding in medium lacking EGF. **d**, Quantification of lumen-containing organoids for each line in the *Foxa1* allelic series. All values of *P* are relative to empty vector control, calculated using unpaired, two-tailed Student's *t*-test. **e**, Histology and immunohistochemistry of organoid lines overexpressing variants of FOXA1 via the doxycycline-inducible pCW vector ten days after seeding. Images from a single biological experiment. H&E, haematoxylin and eosin staining. Scale bars: top, 200 μm; other rows, 100 μm. **f**, Summary of GSEA comparing *Foxa1* wild-type or mutant organoid lines to empty vector control for a basal low (luminal) gene set, the hallmark EMT gene set and a gene set of the top 100 genes induced following ERF knockdown (using short hairpin RNA directed against ERF, shERF) in organoids. Data are from RNA-sequencing (RNA-seq) analysis of three biological replicates for each organoid line. Only comparisons with an FDR of <0.25 are shown with the corresponding normalized enrichment score. Gene sets with a positive normalized enrichment score are enriched in organoids that have either *Foxa1* wild-type or mutant alleles. Data in **a**–**d** are mean ± s.d. Numbers of biological replicates (indicated as dots) and specific *P* values are presented in the source data. $*P < 0.05$, $**P < 0.01$, NS, not significant. All *P* values are relative to overexpression of wild-type FOXA1 unless otherwise noted; unpaired, two-tailed Student's *t*-test.

F266), some of which are predicted to be sites of direct DNA contact[10] (Fig. 1a, Extended Data Fig. 1). Another mutational hotspot is at R219, a DNA-contact site in α-helix 3, which is a highly conserved fold of the FKHD domain that sits in the major groove of target DNA (Extended Data Fig. 1). Finally, 20% of *FOXA1* mutations encode truncations just downstream of the FKHD DNA-binding domain, resulting in the loss of the C-terminal transactivating domain. Annotation of all *FOXA1* mutations in the MSK-IMPACT 504 cohort[11] revealed that Wing2-hotspot mutations—the most common subclass—are found across all disease stages, but are more prevalent in primary locoregional cases (Fig. 1c). There are only four cases of R219 mutation in *FOXA1* in this cohort but, notably, two of these had castration-resistant disease. We therefore expanded the analysis to 1,822 patients by including a larger cohort from MSK-IMPACT and a published cohort from Weill Cornell[12], which is enriched for neuroendocrine prostate cancer (NEPC), and observed significant enrichment ($P < 0.006$) of R219 mutations versus other *FOXA1* mutations in NEPC (3 out of 4) versus adenocarcinoma (8 out of 84) (Fig. 1d).

We next investigated whether *FOXA1* mutation in patients is associated with clinical outcome. In the absence of appropriate longitudinal

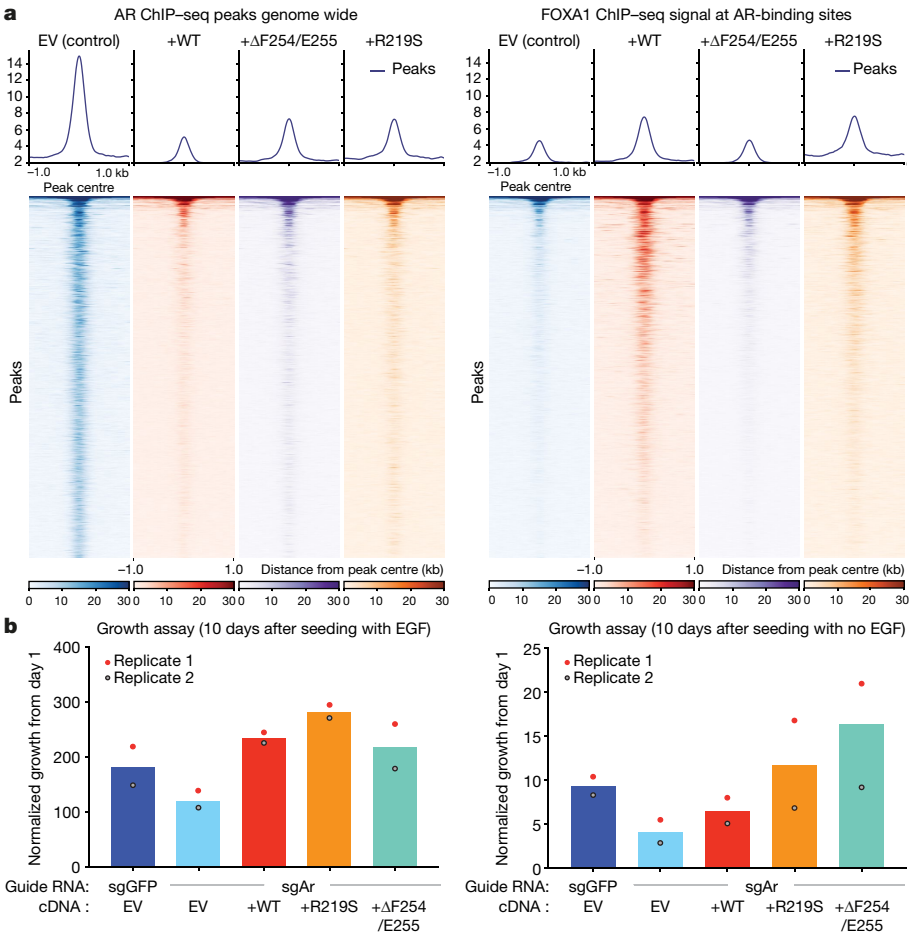


Fig. 3 | FOXA1 expression constricts the AR cistrome and promotes AR-independent growth programs. a, Left, AR ChIP–seq in organoids overexpressing wild-type or mutant FOXA1 compared to control shows significant changes in the AR cistrome in response to FOXA1 expression. Right, FOXA1 ChIP–seq showing FOXA1 binding at same loci. ChIP–seq data are from two biological replicates. Statistical analysis of peaks is shown in Extended Data Fig. 6. **b**, Overexpression of FOXA1 promotes growth in prostate organoids in the setting of significantly reduced AR (CRISPR-mediated depletion in a bulk population), in both standard medium conditions (left) and in the absence of EGF (right). Two independent experiments result in the same growth trends for biological replicates 1 and 2.

data, we generated an RNA signature using mutant *FOXA1* status of The Cancer Genome Atlas (TCGA) samples to interrogate the Decipher GRID cohort of 1,626 primary prostate cancer patients[13] and found that tumours predicted to be *FOXA1* mutant were significantly associated with higher Gleason Scores, shorter time to biochemical recurrence and more rapid progression to metastatic disease than unaltered cases (Extended Data Fig. 1b, c). Together with recent evidence[14], these data suggest that patients with *FOXA1* mutations have a less favourable prognosis.

To characterize a large panel of the most recurrent alterations seen in prostate cancer, including truncating mutations, we generated a FOXA1 reporter construct (Extended Data Fig. 2), and found that all Wing2 mutations, D226N (a mutation in spatial proximity to Wing2 in the protein[10]) and the truncation mutant G275X result in increased transcriptional activity (around twofold) compared to the wild type, whereas mutations at R219 (R219S and R219C) cause impaired activity (around 50% of wild-type activity) (Fig. 2a). To investigate the consequences of *FOXA1* mutations on the growth of prostate cells, we used primary mouse prostate organoid culture (previously used to model tumour initiation)[15] by introducing a series of wild-type or mutant mouse *Foxa1* alleles using doxycycline-inducible lentiviral constructs (Extended Data Fig. 3a–c). Increased expression of wild-type (WT) FOXA1 resulted in a 2–3-fold increase in growth compared to vector control. This relative difference was substantially greater (about 50-fold) after removal of epidermal growth factor (EGF), a critical growth factor for normal organoid proliferation (Fig. 2b). In this setting, nearly all mutants tested showed an increase in growth relative to overexpression of FOXA1(WT), including the two α-helix 3 mutants (R219S and R219C) that had reduced reporter activity, as well as the truncation mutant G275X (Fig. 2c). All 14 mutants promoted growth relative to the empty vector control.

We next examined the histological features of the resulting organoids. We observed that increased expression of FOXA1(WT), FOXA1(D226N) or the Wing2-hotspot mutations all promote exaggerated lumen formation and size (Fig. 2d, e, Extended Data Fig. 3d). By contrast, organoids expressing FOXA1(R219S), and to a lesser extent those expressing FOXA1(R219C), were unable to form measurable lumens, and the bilayer orientation of basal ($p63^+$) and luminal (androgen receptor-positive (AR^+)) cell layers appeared disrupted (Fig. 2e, Extended Data Fig. 3e). This phenotype resembles that of FOXA1-deficient organoids generated using CRISPR–Cas9 (Extended Data Fig. 4a–c), consistent with mouse models[16]. We also repeated the overexpression studies in endogenous-*Foxa1*-deleted organoids using CRISPR-resistant cDNAs encoding two pro-luminal FOXA1 mutants (ΔF254/E255 and D226N) and found that the pro-luminal phenotype was unchanged (Extended Data Fig. 4d–g). Findings from RNA sequencing were consistent with these histologies. Mutants conferring a pro-luminal phenotype showed similarity to ETS-mutant luminal organoids[17] by gene-set enrichment analysis (GSEA), with the notable exception of FOXA1(R219S), which instead showed enrichment of an epithelial–mesenchymal-transition (EMT) program and a repression of the ETS-mutant gene set (Fig. 2f), consistent with

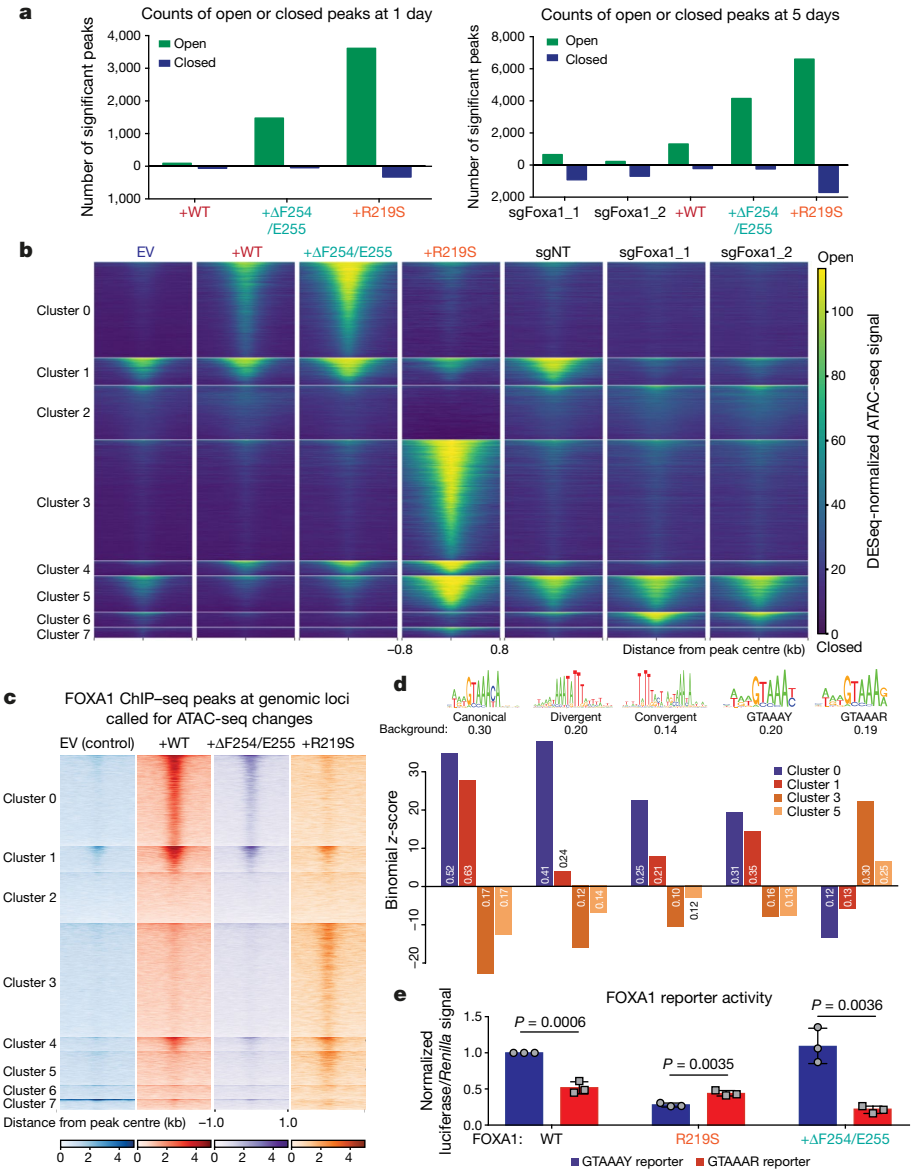


Fig. 4 | *FOXA1* mutations cause marked shifts in the chromatin landscape. a, Number of significant peaks open or closed ($\log_2$(fold change) > 2 for open peaks, $\log_2$(fold change) < -2 for closed peaks) after doxycycline treatment of organoids transfected with pCW-FOXA1 for expression of wild-type or mutant FOXA1 relative to empty vector. Right, includes counts for organoids following CRISPR-mediated *Foxa1* deletion five days after trypsinization, relative to the control sgRNA. Data are from three biological replicates, with FDR <0.05 using two-sided Wald test, with Benjamini–Hochberg FDR correction for multiple observations. **b**, ATAC-seq peak heat maps comparing organoids with (sgFoxa1_1, sgFoxa1_2) or without (sgNT) CRISPR-mediated deletion of *Foxa1* or expression of wild-type or mutant FOXA1 after five days of doxycycline treatment, with eight clusters defined by hierarchical clustering. **c**, FOXA1 ChIP–seq signal at genomic loci matching ATAC-seq clusters defined in **b** shows a similar pattern of peaks correlating FOXA1 binding with open chromatin. Data are from two biological replicates. **d**, Enrichment or depletion of FOXA1 motif variants in clusters that gain accessibility in organoids overexpressing FOXA1(ΔF254/E255) or FOXA1(R219S), including the canonical motif, divergent and convergent dimer motifs, and altered versions of the FOXA1 motif (GTAAAY, similar to canonical, and GTAAAR, non-canonical), expressed as a binomial *Z*-score computed from the number of cluster peaks with >1 motif occurrence relative to background occurrence in all heat map peaks. Occurrence within a given cluster is reported in the bar graph. Positive scores indicate enrichment; negative scores indicate depletion. **e**, Luciferase reporter assay showing activity of FOXA1 variants on GTAAAY (blue) or GTAAAR (red) DNA templates. Luciferase (ratio over *Renilla* luciferase signal, see Methods) signal normalized to signal from FOXA1(WT) on GTAAAY reporter. Data are from three biological replicates, mean ± s.d. Unpaired, two-tailed Student's *t*-test. No significant difference between activity of wild type and R219S on the GTAAAR reporter ($P = 0.2314$). FOXA1(ΔF254/E255) has significantly less activity on GTAAAR than either FOXA1(WT) ($P = 0.0059$) or FOXA1(R219S) ($P = 0.0033$).

its distinct morphology. We also examined the activity of FOXA1 in an in vivo setting[18,19] and observed increased proliferation across all lines, an increase in subcutaneous tumour size in two of four lines (FOXA1(WT) and FOXA1(G275X)), and an increased prevalence of invasive, intraductal basal disease (defined by the loss of AR expression) in tumours derived from FOXA1(R219S) organoids transduced with single-guide RNA (sgRNA) directed against PTEN—consistent with FOXA1(R219S) histology in vitro (Extended Data Fig. 4h–j).

As FOXA1 is a cofactor for AR and *FOXA1* mutant cases in the TCGA cohort have higher AR scores than either normal samples or other subtypes[6], we examined the AR cistrome. The number of AR binding peaks (defined by AR chromatin immunoprecipitation with sequencing (ChIP–seq)) is markedly reduced in organoids overexpressing wild-type or mutant FOXA1 (Fig. 3a, left, Extended Data Fig. 6a). However, FOXA1 binding is enhanced at the sites where AR binding is lost (Fig. 3a, right, $P < 1 \times 10^{-300}$, Extended Data Fig. 5a).

This suggests that FOXA1 may replace AR function at these sites, supported by retention of the increased growth advantage conferred by FOXA1 despite CRISPR-mediated deletion of *Ar* (Fig. 3b, Extended Data Fig. 5b). To reconcile the high AR scores seen in TCGA with this AR-independent growth program, we examined expression levels of the mouse orthologues of the human AR gene signature[20] and found that the majority are induced by FOXA1 (Extended Data Fig. 5c). Thus, while the number of AR-binding sites is substantially reduced, a core set of AR-target genes are maintained in the setting of increased FOXA1 activity. We also investigated whether transcriptomic changes observed in the *Foxa1* mutant mouse organoids were similar to those observed in *FOXA1* mutant human tumours. Remarkably, the human orthologues of differentially expressed genes in FOXA1(ΔF254/E255) mouse organoids were sufficient to cluster *FOXA1* mutant tumours within the TCGA cohort ($P = 2.1 \times 10^{-8}$, Extended Data Fig. 5d).

Given the role of FOXA1 as a pioneer transcription factor, we conducted a genome-wide analysis of changes in open and closed chromatin using ATAC-seq. Expression of FOXA1(WT) led to an increase in open chromatin after five days (more than 1,000 open peaks with significant change in accessibility, false discovery rate (FDR) < 0.05, log fold change of 2 in peak read coverage compared to control) whereas deletion of *Foxa1* led to the opposite, with the closing of around 1,000 peaks. Organoids expressing FOXA1(ΔF254/E255) and FOXA1(R219S) also had increased peak numbers, but these changes occurred substantially faster (in one day) and involved many more peaks (Fig. 4a), consistent with altered pioneering activity.

Unsupervised clustering analysis identified distinct sets of peaks for FOXA1(ΔF254/E255) and FOXA1(R219S) (Fig. 4b). Cluster 0 is largely defined by marked peak changes observed with both FOXA1(WT) and FOXA1(ΔF254/E255), demonstrating that overexpression of wild-type FOXA1 opens new regions of chromatin compared with controls; this effect is amplified in cells expressing FOXA1(ΔF254/E255). By contrast, organoids expressing FOXA1(R219S) gain thousands of distinct peaks (defined by clusters 3 and 5) without changes in cluster 0. ChIP–seq reveals that FOXA1 protein binds at these same ATAC-seq loci (Fig. 4c, Extended Data Fig. 6a–d) and cumulative distributive function plots confirm that there are mutation-specific changes in expression of the genes that map to these newly open chromatin peaks (Extended Data Fig. 6e–h).

Motif analysis revealed enrichment of FOXA1-binding motifs in clusters 0 and 1 (FOXA1(WT) and FOXA1(ΔF254/E255)) (Extended Data Fig. 7a) but not in clusters 3 and 5 (FOXA1(R219S)), despite evidence of FOXA1(R219S)–DNA binding and associated gene-expression changes. However, de novo motif analysis of cluster 3 peaks identified a motif with similarities to the core GTAAA(C/T) FOXA1 binding motif but with substitution of (G/A) for (C/T) at position 6 (Extended Data Fig. 7b). This impression was confirmed by selective enrichment of the (G/A) motif in clusters 3 and 5 versus the (C/T) motif in clusters 0 and 1 (Fig. 4d). To determine whether this motif is functional, we repeated the reporter assays described in Fig. 2a with FOXA1(R219S) and found that FOXA1(R219S) preferentially activates a DNA template modified to reflect the (G/A) bias at position 6, whereas FOXA1(WT) and FOXA1(ΔF254/E255) exhibit substantially higher activity on the canonical (C/T) sequence (Fig. 4e, Extended Data Fig. 7c–e), suggesting a mechanism by which FOXA1(R219S) selectively targets novel genomic loci. Finally, two motifs recently associated with FOXA1 dimers—termed convergent and divergent[21]—were relatively enriched in cluster 0 versus cluster 1, potentially explaining the novel pioneering activity of FOXA1(ΔF254/E255) (Fig. 4d).

Collectively, our analysis of mutant *FOXA1* alleles in prostate cancer reveals unanticipated and diverse consequences for the pioneering function of FOXA1. Wing2 mutants display a gain in pioneering activity that is substantially greater than that observed by overexpression of comparable levels of FOXA1(WT), but both alterations affect nearly identical regions of the genome (cluster 0) that are distinguishable from endogenous FOXA1 sites (cluster 1) on the basis of enrichment of FOXA1 dimer motifs. We postulate that the changes in gene expression associated with these novel open regions contribute to oncogenesis. By contrast, FOXA1 R219 mutants display pioneering function over distinct regions of the genome (clusters 3 and 5) enriched for a variant FOXA1-binding motif that, on the basis of reporter assays, is permissive for the binding of FOXA1 R219 mutants despite mutation of the α-helix 3 consensus DNA-binding residue. Further investigation of the relative DNA-binding affinities of these mutants for the different motifs and the potential role of the Wing2 domain in this retained DNA binding (based on known DNA contacts in the minor groove) is warranted. In both classes of mutations, the biological consequence is lineage plasticity for pro-luminal versus anti-luminal programs.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1318-9.

Received: 10 June 2018; Accepted: 22 May 2019;
Published online 26 June 2019.

1. Pomerantz, M. M. et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
2. Grasso, C. S. et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
3. Gerhardt, J. et al. FOXA1 promotes tumor progression in prostate cancer and represents a novel hallmark of castration-resistant prostate cancer. *Am. J. Pathol.* **180**, 848–861 (2012).
4. Jin, H. J., Zhao, J. C., Ogden, I., Bergan, R. C. & Yu, J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. *Cancer Res.* **73**, 3725–3736 (2013).
5. Barbieri, C. E. et al. Exome sequencing identifies recurrent *SPOP*, *FOXA1* and *MED12* mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
6. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
7. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
8. Annala, M. et al. Frequent mutation of the *FOXA1* untranslated region in prostate cancer. *Commun. Biology* **1**, 122 (2018).
9. Wedge, D. C. et al. Sequencing of prostate cancers identifies new cancer genes, routes of progression and drug targets. *Nat. Genet.* **50**, 682–692 (2018).
10. Ciriello, G. et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163**, 506–519 (2015).
11. Abida, W. et al. Prospective genomic profiling of prostate cancer across disease states reveals germline and somatic alterations that may affect clinical decision making. *JCO Precis. Oncol.* **2017**, https://doi.org/10.1200/PO.17.00029 (2017).
12. Beltran, H. et al. Divergent clonal evolution of castration-resistant neuroendocrine prostate cancer. *Nat. Med.* **22**, 298–305 (2016).
13. Liu, D. et al. Impact of the SPOP mutant subtype on the interpretation of clinical parameters in prostate cancer. *JCO Precis. Oncol.* **2018**, 1–13 (2018).
14. Armenia, J. et al. The long tail of oncogenic drivers in prostate cancer. *Nat. Genet.* **50**, 645–651 (2018).
15. Karthaus, W. R. et al. Identification of multipotent luminal progenitor cells in human prostate organoid cultures. *Cell* **159**, 163–175 (2014).
16. Gao, N. et al. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* **132**, 3431–3443 (2005).
17. Bose, R. et al. ERF mutations reveal a balance of ETS factors controlling prostate oncogenesis. *Nature* **546**, 671–675 (2017).
18. King, J. C. et al. Cooperativity of TMPRSS2–ERG with PI3-kinase pathway activation in prostate oncogenesis. *Nat. Genet.* **41**, 524–526 (2009).
19. Blattner, M. et al. *SPOP* mutation drives prostate tumorigenesis in vivo through coordinate regulation of PI3K/mTOR and AR signaling. *Cancer Cell* **31**, 436–451 (2017).
20. Hieronymus, H. et al. Gene expression signature-based chemical genomic prediction identifies a novel class of HSP90 pathway modulators. *Cancer Cell* **10**, 321–330 (2006).
21. Wang, X. et al. DNA-mediated dimerization on a compact sequence signature controls enhancer engagement and regulation by FOXA1. *Nucleic Acids Res.* **46**, 5470–5486 (2018).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. For all assays except lumen area quantification and histological assessment of xenograft tumors, the investigators were not blinded to allocation during experiments and outcome assessment.

Pan-prostate cancer mutation analysis. The 12 cohorts used for analysis (total of 3,086 samples) included published datasets as well as unpublished data from the MSK-IMPACT 1708 cohort (frozen 25 May 2018), across all stages of prostate cancer (see Supplementary Table 1). Samples were compiled and duplicate samples were pruned to generate a master list of 3,086 prostate cancer cases, which were then stratified on the basis of their *FOXA1* alteration status and the class of mutation in the samples. The Wing2 hotspot includes cases with mutations or indels between H247 and E269. Truncations after the FKHD domain were defined as any frameshift alteration distal to residue E269. Any mutation that did not specifically fall into one of the distinct classes was called 'other'. Sample analysis was performed in part using the CBioPortal for Cancer Genomics[22,23].

3D modelling. Three-dimensional representation of the FKHD domain of FOXA3 complexed with DNA was generated using PyMOL (PDB: 1VTN).

Constructs. To create pCW-Flag-2A-dsRED (pCW-EV), sequences for p2A and DsRED were cloned in the pCW-Cas9 plasmid (Addgene Plasmid #50661) using in-fusion cloning (Takara Bio). To generate pCW-Flag-Foxa1-2A-dsRED (pCW-Foxa1), mouse *Foxa1* cDNA was cloned into pCW-Flag-2A-dsRED using in-fusion cloning (Takara Bio). All primers and sequences are listed in Supplementary Table 2. To generate the sgRNA vector CRISPR-Zeo, GFP from pLKO5.sgRNA.EFS.GFP (a gift from B. Ebert, Addgene plasmid no. 57822) was excised with BamHI and MluI. The zeomycin-resistance gene was removed from lenti sgRNA(MS2)_zeo backbone (a gift from F. Zhang, Addgene plasmid no. 61427) using BsrGI and EcoRI. ZeoR was ligated into the pLKO5.sgRNA.EFS backbone in a four-way ligation using BamHI–BsrGI and EcoRI–MluI adaptors. To create the LVX-UbC-EGFP-Luc2_Hygro construct for visualization of injected cells by live imaging or immunohistochemistry, we first generated the plasmid LVX-UbC-EGFP-Luc2_Puro as follows: 0.72 kb EGFP cDNA from pQCXIP-EGFP[24] was cloned into the BamHI and NotI sites of pLVX-TRE3G-IRES (Clontech, 631362) via a EcoRI–NotI cloning adaptor to make pLVX-TRE3G-EGFP-IRES. The TRE3G promoter was then removed with an XhoI and BamHI digestion, and replaced with the 1.26-kb UbC promoter obtained from Duet011 (Addgene) with a PacI and BamHI digest and using a XhoI–PacI cloning adaptor to make pLVX-UbC-EGFP-IRES. pLVX-UbC-EGFP-Luc2 was then constructed by cloning the 1.7 kb-Luc2 cDNA derived from pGL4.10(luc2) (Promega) with a HindIII and XbaI digest into the MluI and EcoRI sites of pLVX-UbC-EGFP-IRES via MluI–HindIII and XbaI–EcoRI cloning adaptors. The puromycin cassette was replaced with the hygromycin cassette to generate LVX-UbC-EGFP-Luc2_Hygro.

Generation of *FOXA1* mutant cDNA. Site directed mutagenesis was carried out on pCW-Flag-Foxa1-2A-dsRED to induce patient mutations in the cDNA using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent), according to the manufacturer's protocol. Primers were designed using Agilent's QuikChange Primer Design tool (https://www.genomics.agilent.com/primerDesignProgram.jsp). To prevent CRISPR–Cas9 targeting by sgFOXA1_1 sgRNA mutagenesis was used to introduce three silent mutations in the sgRNA recognition sequence (see Extended Data Fig. 8a).

Guide RNA design. Guide RNAs targeting mouse *Foxa1*, *Ar* and *Pten* were generated using the CRISPR design tool (http://crispr.mit.edu). sgFoxa1_1 targets the cDNA near the 5′ end, whereas sg_Foxa14 and sgFoxa1_15 target the FKHD DNA-binding domain. Control guides sgNT (targeting safe harbour locus AAVS1[25]) and sgGFP were used. All guide RNAs were cloned into lentiCRISPRv2 (Addgene no. 52961), lentiGuide-Puro (Addgene no. 52963) or CRISPR-Zeo using BsmbI digest, according to the Zhang laboratory protocol. For cells carrying CRISPR-Zeo or lentiGuide-Puro, lentiCas9-Blast (a gift from F. Zhang, Addgene plasmid no. 52962) was used as the Cas9 source.

FOXA1 luciferase reporter pGL-6xFBS-Luc. Oligonucleotide fragments containing six tandem FKHD-consensus (canonical or non-canonical) motifs with 5-bp spacers (Supplementary Table 2) were cloned into pGL4.28 luc2CP/minP/hygro (Pomega) between HindIII and XhoI restriction sites. Oligonucleotide sequences were verified using Sanger sequencing. Canonical FOXA1-binding sites were based on the top binding motifs predicted on the basis of ChIP–seq results in HepG2 cells[26], whereas the non-canonical FOXA1-binding site was based on the top hit of de novo motif analysis of ATAC-seq cluster 3 using HOMER (Extended Data Fig. 10). The pGL-5xFBS-Luc was transiently transfected using Lipofectamine 2000 (Thermo Fisher) into lentiX293T cells (Clonetech) along with CMV-Renilla (pRL-CMV *Renilla*, Promega) as an internal control. Response ratios are expressed relative to signal obtained for the positive-control wells transfected with 170 ng of pCMV6-mFOXA1mycDDK (Origene MR225487), which was set to 1, and the negative-control well receiving 170 ng of 'stuffer' DNA (pCW-Flag-2A-dsRED (pCW-EV), no exogenous FOXA1), which was set to 0. To test the response of these reporters to varying levels of FOXA1 introduced into the system, ratios of pCMV6-mFOXA1mycDDK and pCW-EV constructs were altered, keeping the total amount of DNA transfected into each well constant. In evaluating the relative response ratios between FOXA1(WT) and various mutants, one concentration of cDNA (170 ng per well) was used and relative response ratios reflect activity of given variant on the reporter. Luminescence measurements were taken 24 h after transfection. All results are means and standard deviations from experiments performed in biological triplicates (as a minimum, *n* ranging from 3 to 7) (see source data for Fig. 2 for details), and Firefly luciferase activity of individual wells was normalized against *Renilla* luciferase activity.

Organoid lines. All parental organoid lines were established in our laboratory as previously described[15]. The blue red organoids (BRO) line was established from mice harbouring red fluorescent protein (RFP) driven by a composite human keratin 18 promoter and cerulean fluorescent protein (CFP) driven by a bovine keratin 5 promoter[27]. BROs were transduced with lentiCrisprv2 carrying either sgNT or sgFoxa1_1 and selected using puromycin. BRO lines were maintained in standard mouse organoid media conditions[15]. K14-1 organoids were derived from mice harbouring an actin–GFP fusion protein driven by a human keratin 14 promoter[28]. K14-1 organoids were transduced with the allelic series of pCW-Foxa1 wild-type or mutant constructs, as well as pCW-EV as a control. Bulk cells were selected using puromycin. K14-1 organoids were maintained in standard mouse organoid media conditions[15], with 2.5 ng/ml EGF supplementation. For rescue experiments of either *Foxa1* or *Ar* deletion, K14-1 organoids carrying pCW-Foxa1 constructs were subsequently transduced with lentiCas9-Blast, bulk selected with blasticidin, and then transduced with either CRISPR-Zeo-sgFoxa1_1 or sgGFP, or sgAr and bulk selected with zeocin. Rosa26-Cas9-sgPTEN-luc2-pCW-FOXA1 organoids were derived from a homozygous Rosa26 Lox-stop-Lox Cas9 mouse (C67BL/6J background, Jackson Laboratory 026175) and transduced with adenoCre-GFP in vitro for expression of Cas9. These cells were then transduced with lentiGuide-Puro-sgPten and bulk selected with puromycin, transduced with LVX-UbC-EGFP-Luc2_Hygro and bulk selected with hygromycin, transduced with the allelic series of pCW-Foxa1 wild-type or mutant constructs or pCW-ERG, as well as pCW-EV as a control, and sorted for dsRED expression to enrich for transduced cells.

Organoid culture. Mouse organoids were sorted, cultured in 3D and transduced with lentiviruses as described previously[15,29]. Organoids infected with pCW-EV, pCW-FOXA1, or LentiCrispV2 constructs were selected with 2 μg/ml puromycin for 5 days, 3–4 days after transduction, while those infected with CRISPR-Zeo were selected with 30 μg/ml for 7 days, 3–4 days after transduction. Transduction with Lenti-Cas9-Blast was followed by five days of selection in 10 μg/ml blasticidin. Preparation of 3D organoids for histology was carried out as previously described[15]. Haematoxylin and eosin staining and immunohistochemistry was carried out by the MSKCC Molecular Cytology Core. Cells were confirmed to be free of mycoplasma using the Lonza MycoAlert Mycoplasma Detection Kit (LT07-318).

Growth assays. Organoids were treated with doxycycline (500 ng/ml) to induce expression of the FOXA1-2A-DsRED fusion, then sorted two days later to enrich for DsRED$^+$ cells. Cells were seeded at a density of 100 cells per μl (2,000 cells per 20-μl dome, three domes per line per time point, each dome in a single well of a 48-well plate) and maintained on doxycycline for the duration of the assay, refreshing media every 2–3 days. Y-27632 was supplemented for the first feeding at 10 μM. To measure proliferation, matrigel domes were washed with PBS, and then resuspended in 100 μl of PBS, and CellTiter-Glo 2.0 Assay (Promega) was used, following the manufacturer's instructions. Triplicate values for each time point were averaged, and all values on subsequent days were normalized to the day 1 reading. Experiments were repeated at least three times independently and each line was normalized to the empty vector control readings for a given replicate.

Lumen formation assays. Organoids were treated with doxycycline (500 ng/ml) to induce expression of the FOXA1-2A-DsRED fusion. Doxycycline-treated cells were sorted two days later to enrich for DsRED$^+$ cells. Sorted cells were seeded in matrigel at a density of 3 cells per μl (eight 25-μl domes per organoid line) and maintained on doxycycline for the duration of the assay, with fresh medium every 2–3 days. Y-27632 was supplemented for the first feeding at 10 μM. After 10 days, organoids were scored for the presence or absence of a visible lumen by bright-field microscopy, and the percentage of the total number of organoids that possessed a lumen was determined from examining ~50–200 organoids in a typical experiment. In CRISPR organoid lines, sorting was not performed for the lumen formation assay. Instead cells were trypsinized to a single-cell suspension, counted using trypan blue exclusion, and then seeded as described above. Experiments were repeated three times independently.

Lumen area measurements. Organoids were treated with doxycycline (500 ng/ml) to induce expression of the FOXA1-2A-DsRED fusion. Doxycycline-treated cells were sorted two days later to enrich for DsRED$^+$ cells. Sorted cells were seeded in matrigel in a dilution series of densities ranging from 32 cells per μl down to 4 cells per μl (five domes per density per line) and maintained on doxycycline for the duration of the assay, with the medium refreshed every 2–3 days. Y-27632 was

supplemented for the first feeding at 10 μM. After 10 days, the area of each visible lumen was measured using light microscopy and Nikon NIS Elements software. In a typical experiment, ~30–50 organoids were measured by an observer blinded to organoid genotypes.

Western blot. Membranes were probed with antibodies directed against AR (1:1,000, ER179(2), Abcam), FOXA1 (1:1,000, Ab2, Sigma), cyclophilin B (1:1,000, EPR12703(B), Abcam), Flag (1:1,000, M2, Sigma) or PTEN (1:1,000, D4.3, Cell Signaling). Signal was visualized with secondary HRP-conjugated antibodies and chemiluminescent detection.

Immunohistochemistry. Organoids and tumours were processed and stained as described previously[15]. The following antibodies were used for staining on mouse organoids and organoid-derived xenografts: HNF-3-α/FOXA1 antibody (5 μg/ml, 3B3NB, Novus Biologicals), AR (1:1,000, N-20, Santa Cruz), p63 (1:800, 4A4, Ventana). Staining was visualized with BrightVision (Dako), Ki67 (Abcam ab15580 at 1 μg/ml).

In vivo experiments. In vivo xenograft experiments were performed by subcutaneous injection of 2×10^6 dissociated organoid cells (Rosa26-Cas9-sgPTEN-luc2-pCW-FOXA1 or ERG) resuspended in 100 μl of 50% matrigel (BD Biosciences) and 50% growth medium into the flanks of five 8-to-12-week-old male NOD. Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ mice (005557, The Jackson Laboratory) to yield ten tumours per group. As soon as palpable, tumour volume was measured weekly using the tumour-measuring system Peira TM900 (Peira). Tumours were then collected at given time points for histology using 4% paraformaldehyde. Histological assessment was carried out by an observer blinded to tumour genotypes. All animal experiments were performed in compliance with the guidelines of the Research Animal Resource Center of the Memorial Sloan Kettering Cancer Center. In accordance with our IACUC and our approved protocol, none of the mice exceeded the maximum tumour burden allowed (total for both sides) of 2,000 mm^3.

RNA isolation and sequencing. RNA was extracted from organoids using an RNeasy Kit (Qiagen). Freshly sorted dsRED$^+$ cells were seeded in triplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected at the appropriate time points. Library preparation and sequencing were performed by the New York Genome Center, where RNA-seq libraries were prepared using the TruSeq Stranded mRNA Library Preparation Kit (Illumina) in accordance with the manufacturer's instructions. In brief, 100 ng of total RNA was used for purification and fragmentation of mRNA. Purified mRNA underwent first and second strand cDNA synthesis. cDNA was then adenylated, ligated to Illumina sequencing adapters, and amplified by PCR (using 10 cycles). Final libraries were evaluated using fluorescent-based assays including PicoGreen (Life Technologies) or Qubit Fluorometer (Invitrogen) and Fragment Analyzer (Advanced Analytics) or BioAnalyzer (Agilent 2100), and were sequenced on an Illumina HiSeq2500 sequencer (v.4 chemistry, v.2 chemistry for Rapid Run) using 2×50-bp cycles. Reads were aligned to the mm10 mouse reference genome using STARaligner[30] (v.2.4.2a). Quantification of genes annotated in Gencode vM2 was performed using featureCounts (v.1.4.3) and transcripts were quantified using Kalisto[31]. Quality control statistics were collected with Picard (v.1.83) and RSeQC[32] (http://broadinstitute.github.io/picard/). Normalization of feature counts was done using the DESeq2 package[33].

Analysis of RNA-seq from mouse organoids and patient samples. The gene read count data of TCGA primary prostate cancer were downloaded using the GDC tool. The mouse and human homologous genes were downloaded from Mouse Genome Informatics from The Jackson Laboratory (http://www.informatics.jax.org/homology.shtml). Differential expression analyses were performed using DESeq2 (https://bioconductor.org/packages/release/bioc/html/DESeq2.html) based on the gene read count data. Multiple-hypothesis testing was considered by using Benjamini–Hochberg correction. The statistical significance of the overlap between two groups of genes was tested using Fisher's exact test. GSEA was performed using the JAVA program (http://www.broadinstitute.org/gsea) and run in pre-ranked mode to identify enriched signatures. The GSEA plot, normalized enrichment score and FDR and q values were derived from GSEA output. The following gene sets were used: Hallmark gene sets, Neuroendocrine high[12], Basal low[34], and shERF up[17].

Prostate cancer tumour samples and microarray data. A total of 1,959 radical prostatectomy tumour expression profiles were used for training and testing. For training and testing, we used RNA-seq expression and DNA mutation data from TCGA prostate cancer project[6] ($n = 333$). For testing, the expression profiles of retrospective ($n = 1{,}626$) were derived from the Decipher GRID registry (NCT02609269). The retrospective GRID cohort was pooled from seven published microarray studies: Cleveland Clinic[35] (CCF), Erasmus MC[36], Johns Hopkins[37] (JHMI), Memorial Sloan Kettering[38] (MSKCC), Mayo Clinic[39,40] (Mayo I and Mayo II), and Thomas Jefferson University[41] (TJU). Associated accession numbers are: GSE79957, GSE72291, GSE62667, GSE62116, GSE46691, GSE41408, and GSE21032. DNA and RNA from the TCGA cohort were extracted from fresh frozen radical prostatectomy tumour tissue, as previously described[6]. RNA from the GRID cohorts was extracted from routine formalin-fixed, paraffin embedded (FFPE) radical prostatectomy tumour tissues, amplified and hybridized to Human Exon 1.0 ST microarrays (Thermo Fisher).

***FOXA1* mutant transcriptional signature.** By following the similar strategy as previously reported for *SPOP* mutants[13], we developed the *FOXA1* mutant transcriptional signature that includes 67 genes differentially expressed between *FOXA1* mutant and wild-type samples from TCGA prostate cancer RNA-seq data. The low-expressed genes (mean RNA-seq by expectation maximization (RSEM) <1) were filtered before the analysis. Specifically, we identified significantly differentially expressed genes by comparing *FOXA1* mutants within FKHD DNA-binding domain and wild-type cases as determined from DNA mutational analyses among TCGA samples lacking *ETS* family gene fusions (*ERG*, *ETV1*, *ETV4* and *FLI1*), using Wilcoxon rank-sum test and controlled for false discovery using Benjamini–Hochberg adjustment (FDR ≤ 0.05).

SCaPT development based on *FOXA1* mutant transcriptional signature and SVM model. To predict tumours in the *FOXA1* mutant subclass in the absence of DNA sequencing data (that is, microarray datasets), we developed the subclass predictor based on transcriptional data (SCaPT) model based on the support vector machine (SVM) model. Given a set of training data marked with two categories, SVM builds a model that assigns testing data into one category or the other, making it a non-probabilistic binary linear classifier. In our SCaPT model, the training data were defined as the transcriptional scores of *FOXA1* mutant signature from TCGA cohort. The testing data would be the transcriptional z scores from RNA-seq or microarray expression data of *FOXA1* mutant signature.

Prostate cancer molecular subclass prediction by decision tree. In each individual study of retrospective and prospective GRID cohorts, the *FOXA1* mutant subclass was first predicted using the SCaPT model. Next, using a decision tree and previously developed microarray-based classifiers for the ERG$^+$ and ETS$^+$ subtypes, we classified the remaining cases in each cohort. Some cases with both predicted *FOXA1* mutant and ERG$^+$ETS$^+$ status were classified as conflict subclass, and the rest without *FOXA1* mutant calling and outlier expression were considered as 'other' subclass.

Statistical analysis of human data. Statistical analyses were performed in R v.3.4.0 (R Foundation). All statistical tests were two-sided with a significance level of $P < 0.05$. Univariate logistic regression analyses were performed on the combined cohort to test the statistical association between *FOXA1* mutant status and clinical variables, including age, race, pre-operative prostate-specific antigen (PSA), Gleason score, lymph node invasion (LNI), surgical margin status (SMS), extracapsular extension (ECE) and seminal vesicle invasion (SVI). We evaluated the associations between *FOXA1* mutant status and patient outcomes including biochemical recurrence (BCR), metastasis (MET) and prostate cancer specific mortality (PCSM) on the basis of Kaplan–Meier analysis.

ATAC-seq. Freshly sorted cells carrying pCW constructs (dsRED$^+$) were seeded in triplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected at the appropriate time points. CRISPR cell lines carried LentiCRISPRv2 with either the control guide (sgNT), guide 14 for FOXA1 (sgFOXA1_1) or guide 15 for FOXA1 (sgFOXA1_2). At time of collection, cells were trypsinized, and 50,000 cells (counted by using trypan blue exclusion) were processed for ATAC-seq as follows. After a wash step in cold cell wash buffer (CWB; 10 mM Tris-HCl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$), outer membranes were disrupted in lysis buffer (CWB + 0.1% NP40) for 2 min on ice. The lysis reaction was stopped with the addition of 1 ml CWB. After a centrifugation step at 1,500*g* for 10 min, pelleted nuclei were kept for the next step. In a 50-μl final volume, tagmentation was performed for 30 min at 37 °C, using the Nextera DNA library prep kit (Illumina FC-121-1030). After addition of SDS to 0.2% final concentration, DNA was purified on AMPure XP beads (Beckman Coulter A63881) using a 2:1 ratio (v/v) of beads:tagmented DNA. Freshly eluted DNA was barcoded and amplified in 110-μl PCR volume (NEB Next Q5 Hot Start HiFi PCR, M0543L) to generate a library with the following PCR program: 65 °C, 5 min; 98 °C, 30 s; 11 cycles of (98 °C, 10 s ramping to 65 °C, 30 s), 4 °C hold. Quality control of the libraries was performed with a Bioanalyzer 2200 (Agilent Technologies, D1000 screentapes and reagents, 5067-5582) to assess size range of amplified DNA fragments and with Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher P11496) to quantify the DNA fragments generated. ATAC libraries were then pooled at equimolar concentration and were sequenced multiplexed on the Illumina HiSeq with 50-bp paired-ends.

ATAC-seq data and preprocessing. ATAC-seq data preprocessing was performed as previously described. Raw ATAC-seq reads were trimmed and filtered for quality using Trim Galore! v.0.4.5 (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) powered by CutAdapt v.1.16 (https://doi-org.proxy.library.cornell.edu/10.14806/ej.17.1.200) and FastQC v.0.11.7 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Paired end reads were aligned to the mm10 genome using Bowtie2 v.2.3.4.1 in very sensitive local mode (-q –local –very-sensitive-local –no-discordant –no-mixed –dovetail -I 10 X 20), and paired reads that mapped

to different chromosomes or that mapped too far away were discarded. Unpaired reads, discordant reads, reads with mapQ <20, or SAM flags 0x4 and 0x400, as well as reads marked as optical or PCR duplicates using picard MarkDuplicates v.2.18.3-SNAPSHOT and reads overlapping the ENCODE mm10 functional genomics regions blacklist (at mitra.stanford.edu/kundaje/akundaje/release/blacklists/mm10-mouse/mm10.blacklist.bed.gz) were removed to improve the quality of the retained fragments. To correct for the fact that the Tn5 transposase binds as a dimer and inserts two adapters in the Tn5 tagmentation step, all positive-strand reads were shifted 4 bp downstream and all negative-strand reads were shifted 5 bp upstream to centre the reads on the transposase-binding event.

Overall mapping statistics confirmed high-quality ATAC-seq data, with a high alignment rate (over 76.8% in all samples) and high coverage (over 30 million aligned read pairs per sample) across experiments (Supplementary Table 13). As an additional quality control metric, we confirmed that all ATAC-seq libraries displayed the expected insert size distribution computed from aligned read pairs, with nucleosome-free, mono-nucleosomal, and di-nucleosomal modes (see Extended Data Fig. 8a for representative plots).

ATAC peak calling, reproducibility analysis and atlas creation. We then pooled the shifted reads by sample and identified peaks using MACS2 with a threshold of FDR-corrected $P < 0.01$ using the Benjamini–Hochberg procedure for multiple-hypothesis correction. As called peaks may be caused by noise in the assay and not reflect true chromatin accessibility, we calculated an irreproducible discovery rate (IDR) for all pairs of replicates across a cell type. Given two ranked lists of events from replicate experiments, in this case peak calls ranked by *P* value, IDR estimates a threshold at which events are no longer reproducible. Using this measure, we excluded peaks that were not reproducible (IDR <0.005) in at least one pair of replicates for at least one cell type or time point.

Reproducible peaks from each cell type were combined to create a genome-wide atlas of accessible chromatin regions. Reproducible peaks from different samples were merged if they overlapped by more than 75%. This produced an atlas of ~182,800 reproducible peaks of median width 586 bp. The numbers of reproducible peaks per time point and organoid line are provided in Supplementary Table 14. Track diagrams at specific loci visually confirm that replicate ATAC-seq experiments show reproducible accessible sites (Extended Data Fig. 8b).

Assignment of ATAC-seq peaks to genes. The RefSeq transcript annotations of the mm10 mouse genome were used to define the genomic location of transcription units. For genes with multiple gene models, the longest transcription unit was used for the gene locus definition. ATAC-seq peaks located in the body of the transcription unit, together with the 2-kb regions upstream of the transcription start site and downstream of the 3′ end, were assigned to the gene. If a peak was found in the overlap of the transcription units of two genes, one of the genes was chosen arbitrarily. Intergenic peaks were assigned to the gene with a transcription start site or 3′ end that was closest to the peak. In this way, each peak was unambiguously assigned to one gene. Peaks were annotated as promoter peaks if they were within 2 kb of a transcription start site. Non-promoter peaks were annotated as intergenic, intronic or exonic according to the relevant RefSeq transcript annotation. The atlas-wide distribution of promoter, intergenic or exonic peak assignment was consistent with high-quality ATAC-seq datasets (Extended Data Fig. 9), with 31.6% of peaks at promoters and the rest nearly equally divided between intergenic and intronic regions, with a small fraction annotated as exonic.

Differential peak accessibility. Reads aligning to the atlas peak regions were counted using htseq-count (-r pos s no). Differential accessibility of the peaks was assessed by applying DESeq2 v.1.18.1 to this count table, considering all pairwise comparisons of cell types. Peaks were defined as differentially accessible if they satisfied an FDR-corrected $P < 0.05$ and if the magnitude of the DESeq-normalized counts changed by a stringent factor of 4 or more between at least one pairwise comparison of organoid line to control (the comparisons used were EV day 1 vs FE255 day 1, EV day 1 vs R219 day 1, EV day 1 vs WT day 1, EV day 5 vs FE255 day 5, EV day 5 vs R219 day 5, EV day 5 vs WT day 5, WT day 1 vs FE255 day 1, WT day 1 vs R219 day 1, EV day 5 vs FE255 day 5, WT day 5 vs R219 day 5, sgNT day 5 vs sgFOXA1-sg1 day 5, and sgNT day 5 vs sgFOXA1-sg2 day 5) (two-sided Wald test, with Benjamini–Hochberg correction for multiple observations). MA plots for pairwise differential accessibility analyses confirmed that normalization was appropriate and that differential peaks displayed robust changes (see Extended Data Fig. 10 for representative plots and Supplementary Table 16 for numbers of differentially accessible peaks). These analyses produced a set of ~20,500 differentially accessible peaks of median width 410 bp; as expected, differential peaks were enriched for intergenic and intronic annotations and depleted for promoter annotations (Extended Data Fig. 9).

ATAC-seq peak clustering. The ATAC-seq peak heat maps were created using the DESeq size-factor normalized read counts, applying the variance-stabilizing transformation to the full peak atlas, selecting the differentially accessible peaks, and then clustering using hierarchical clustering with the ward.D distance metric. Clusters were defined by cutting the hierarchical clustering at the first 8 bifurcations of the dendrogram by ward.D distance. The number of clusters was chosen to be eight based on observation of biologically interesting patterns of accessibility, and then peaks were sorted within each cluster by maximum signal.

Peak heat maps. Heat maps (tornado plots) of peaks were generated by combining signals across replicates and binning the region ±750 bp around the peak summit in 1-bp bins after adjusting the reads for Tn5-induced bias, resulting in one signal track for each cell type or time point. Heat maps were generated using deeptools v.3.0.2.

De novo transcription factor motif analysis. The Homer v.4.10 utility findMotifsGenome.pl was used to identify the top ten transcription factor motifs enriched in each of the clusters produced by deeptools from each time point relative to genomic background. The top motifs were reported and compared to the Homer database of known motifs and then manually curated to restrict to transcription factors that are expressed based on RNA-seq data and to group similar motifs from transcription factors belonging to the same family.

FIMO motif search. Motif enrichment was performed relative to the 8 clusters defined by hierarchical clustering of 20,523 differentially accessible peaks (described above). Each ATAC-seq peak in the atlas was scanned for 718 transcription factor motifs in the *Mus musculus* CIS-BP database[42] using FIMO[43] of MEME suite[44], using the default *P* value cut-off of 1×10^{-4}. The background sequence distribution for motif analysis was based on nucleotide frequencies in the full set of 20,523 differentially accessible peaks (A = T = 0.2711, C = G = 0.2289). Of the 718 motifs in the database, 713 had a match within at least one peak among the differentially accessible peaks.

FIMO motif analysis. We restricted the analysis to 298 transcription factors that had a median RNA-seq expression across biological replicates of above 5 reads per kilobase of transcript per million mapped reads in at least one organoid line or time point. In addition, CTCF and CTCFL, DNA-binding proteins associated with 3D chromatin structure, were excluded. To rank the level of enrichment of transcription factor motifs in each cluster relative to the background, the number of peaks containing each motif was calculated for each cluster and for the full set of differentially accessible peaks. Enrichment–depletion scores for each motif in a cluster were reported as binomial *Z*-scores relative to the background of motif occurrences in the set of differential ATAC-seq peaks. Namely, if p represents the probability that a peak in the background set contains an occurrence of the motif, then the binomial *Z*-score for a cluster of size N with C peaks containing the motif is $\frac{C - Np}{\sqrt{Np(1-p)}}$. While these *Z*-scores do not incorporate a correction for multiple hypotheses, in practice the top-ranked motifs have such strong enrichments that they would still be highly significant after correction.

Non-canonical FOXA1 motif analysis. To examine enrichment–depletion of non-canonical FOXA1 motifs, we considered four additional motifs. First, we examined previously reported convergent and divergent FOXA1 dimer motifs. Second, we altered the canonical FOXA1 motif by replacing position 6 of the core GTAAAC/T pattern with either and equal probability of C/T (similar to canonical) or an equal probability of A/G (non-canonical). We used FIMO to search for hits of these motifs across differential peaks and reported enrichment–depletion within clusters as binomial *Z*-scores as before.

ChIP–seq. Freshly sorted cells carrying pCW constructs (dsRED$^+$) were seeded in duplicate per infected construct at the start of the assay, and moving forward, replicates were processed independently, collected following five days of doxycycline treatment. At time of collection, cells were trypsinized and 70,000 cells (counted by using trypan blue exclusion) were processed for ChIP–seq as follows. Cells were fixed with formaldehyde (1%) and the reaction was quenched with glycine 1.25 M and Tris 1 M pH 8. Fixed cells were lysed with SDS lysis solution containing protease inhibitors. Re-suspended pellets were sonicated, precipitated with antibodies (HNF-3 alpha/FoxA1 antibody (3B3NB) (Novus Biologicals), AR (ER179(2), Abcam) and protein A/G bead complex. The chromatin and immune complex were sequentially washed with a low-salt solution, high-salt solution, LiCl solution and Tris-NaCl solution. Chromatin was eluted from the complex with a solution containing 1% of SDS and 0.1 M $NaHCO_3$. Cross-linking between DNA and protein was reversed by adding NaCl solution and incubating at 65 °C overnight. Libraries were made using NEBNext Ultra II DNA library prep kit for Illumina (NEB E7645L). Quality control was performed with Bioanalyzer 2200 (Agilent Technologies, D1000 screentapes and reagents, 5067-5582) to assess size range of amplified DNA fragments, and with Quant-iT PicoGreen dsDNA Assay Kit (Thermo Fisher P11496) to quantify the DNA fragments generated. ChIP libraries were then pooled at equimolar concentration and were sequenced multiplexed on the Illumina HiSeq with 50-bp paired-end sequencing.

Bioinformatic analysis of ChIP–seq. Raw reads were first trimmed with Trimmomatic[45] (v.0.35, options: LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36) to remove adapters and low-quality sequences. They were then aligned with Bowtie2[46] (v.2.2.6, options:–local–mm–no-mixed–no-discordant) using mm10 genome. After alignment, PCR duplicates were removed with Picard

tools (http://broadinstitute.github.io/picard/) (MarkDuplicates v.2.9.0) and peaks were called individually for each replicate with MACS2[47] (v.2.1.0.20151222,–options: keep-dup 1 -g mm -p 0.05). These called peaks between replicates were then used with IDR[48] (v.2.0.2) framework to identify reproducible peaks. Deeptools (v.3.1.3) was used for visualization and HOMER (v.4.10.3) was used for discovering de novo motifs.

ChIP–seq normalization and analysis. To analyse ChIP–seq signal for AR and FOXA1 in each organoid line relative to ATAC-seq clusters, we normalized ChIP–seq data across experiments based on background signal, namely by defining flanking regions of reproducible peaks and using DEseq scaling factors relative to these regions for library size normalization. To compare AR or FOXA1 binding between a pair of organoid lines with respect to an ATAC-seq cluster, we compared the corresponding distributions of normalized ChIP–seq signal over peaks in the cluster by a one-sided Wilcoxon rank-sum test.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The described RNA-seq, ATAC-seq and ChIP–seq data have been deposited in the Gene Expression Omnibus under the following accession numbers: GSE128667 (all data), GSE128421 (ATAC-seq sub-series), GSE128666 (RNA-seq subseries) and GSE128867 (ChIP–seq subseries). Patient predicted *FOXA1* mutant status and outcome data from Decipher GRID are available from the authors upon reasonable request.

22. Gao, J. et al. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* **6**, pl1 (2013).
23. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
24. Watson, P. A. et al. Constitutively active androgen receptor splice variants expressed in castration-resistant prostate cancer require full-length androgen receptor. *Proc. Natl Acad. Sci. USA* **107**, 16759–16765 (2010).
25. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR–Cas9 system. *Science* **343**, 80–84 (2014).
26. Motallebipour, M. et al. Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP–seq. *Genome Biol.* **10**, R129 (2009).
27. Peng, W., Bao, Y. & Sawicki, J. A. Epithelial cell-targeted transgene expression enables isolation of cyan fluorescent protein (CFP)-expressing prostate stem/progenitor cells. *Transgenic Res.* **20**, 1073–1086 (2011).
28. Vaezi, A., Bauer, C., Vasioukhin, V. & Fuchs, E. Actin cable dynamics and Rho/Rock orchestrate a polarized cytoskeletal architecture in the early steps of assembling a stratified epithelium. *Dev. Cell* **3**, 367–381 (2002).
29. Koo, B.-K. et al. Controlled gene expression in primary *Lgr5* organoid cultures. *Nat. Methods* **9**, 81–83 (2011).
30. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
31. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
32. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185 (2012).
33. Love, M. I., Huber, W., & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
34. Smith, B. A. et al. A basal stem cell signature identifies aggressive prostate cancer phenotypes. *Proc. Natl Acad. Sci. USA* **112**, E6544–E6552 (2015).
35. Klein, E. A. et al. A genomic classifier improves prediction of metastatic disease within 5 years after surgery in node-negative high-risk prostate cancer patients managed by radical prostatectomy without adjuvant therapy. *Eur. Urol.* **67**, 778–786 (2015).
36. Boormans, J. L. et al. Identification of *TDRD1* as a direct target gene of *ERG* in primary prostate cancer. *Int. J. Cancer* **133**, 335–345 (2013).
37. Ross, A. E. et al. Tissue-based genomics augments post-prostatectomy risk stratification in a natural history cohort of intermediate- and high-risk men. *Eur. Urol.* **69**, 157–165 (2016).
38. Taylor, B. S. et al. Integrative genomic profiling of human prostate cancer. *Cancer Cell* **18**, 11–22 (2010).
39. Erho, N. et al. Discovery and validation of a prostate cancer genomic classifier that predicts early metastasis following radical prostatectomy. *PLoS ONE* **8**, e66855 (2013).
40. Karnes, R. J. et al. Validation of a genomic classifier that predicts metastasis following radical prostatectomy in an at risk patient population. *J. Urol.* **190**, 2047–2053 (2013).
41. Den, R. B. et al. Genomic prostate cancer classifier predicts biochemical failure and metastases in patients after postoperative radiation therapy. *Int. J. Radiat. Oncol. Biol. Phys.* **89**, 1038–1046 (2014).
42. Weirauch, M. T. et al. Determination and inference of eukaryotic transcription factor sequence specificity. *Cell* **158**, 1431–1443 (2014).
43. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27**, 1017–1018 (2011).
44. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
46. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
47. Feng, J., Liu, T., Qin, B., Zhang, Y. & Liu, X. S. Identifying ChIP–seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740 (2012).
48. Li, Q., Brown, J. B., Huang, H. & Bickel, P. J. Measuring reproducibility of high-throughput experiments. *Ann. Appl. Stat.* **5**, 1752–1779 (2011).

Acknowledgements We thank P. Iaquinta, B. Carver, Z. Cao, I. Ostrovnaya, H. Hieronymous, W. Abida, E. Wasmuth, K. Lawrence, T. Nadkarni, S. P. Gao and all members of the Sawyers laboratory for comments; Memorial Sloan Kettering Cancer Center core facilities; N. Fan and D. Yarilin from the MSKCC Molecular Cytology Core Facility; the MSKCC Integrated Genomics Operation; the New York Genome Center for RNA sequencing; and R. J. Karnes (Department of Urology, Mayo Clinic), R. B. Den (Department of Radiation Oncology, Thomas Jefferson University), E. A. Klein (Glickman Urological and Kidney Institute, Cleveland Clinic) and Bruce Trock (Department of Urology, Johns Hopkins University) for providing access to patient outcome data. Some of the results shown here are in part based on data generated by the TCGA Research Network (https://www.cancer.gov/tcga). E.J.A. was supported by an American Association for Cancer Research Basic Cancer Research Fellowship, the MSKCC Translational Research Oncology Training Program and the MSKCC Functional Genomics Initiative. R.D. was supported by NIH training grant 1T32GM083937. Z.Z. is supported by the NCI Predoctoral to Postdoctoral Fellow Transition Award (F99/K00 award ID: F99CA223063). R.B. was supported by grants from the Department of Defense (W81XWH1510277), NCI (1K08CA226348-01) and the Prostate Cancer Foundation. D.L., A.S. and C.E.B. were supported by the NCI (K08CA187417-01, R01CA215040-01 and P50CA211024-01 to C.E.B.), a Urology Care Foundation Rising Star in Urology Research Award (C.E.B.), a Damon Runyon Cancer Research Foundation MetLife Foundation Family Clinical Investigator Award (C.E.B.) and the Prostate Cancer Foundation (C.E.B). C.L.S. is an investigator of the Howard Hughes Medical Institute and this project was supported by NIH grants CA155169, CA193837, CA224079, CA092629, CA160001 and CA008748, the Starr Cancer Consortium grant I10-0062 and the Functional Genomics Initiative at MSKCC.

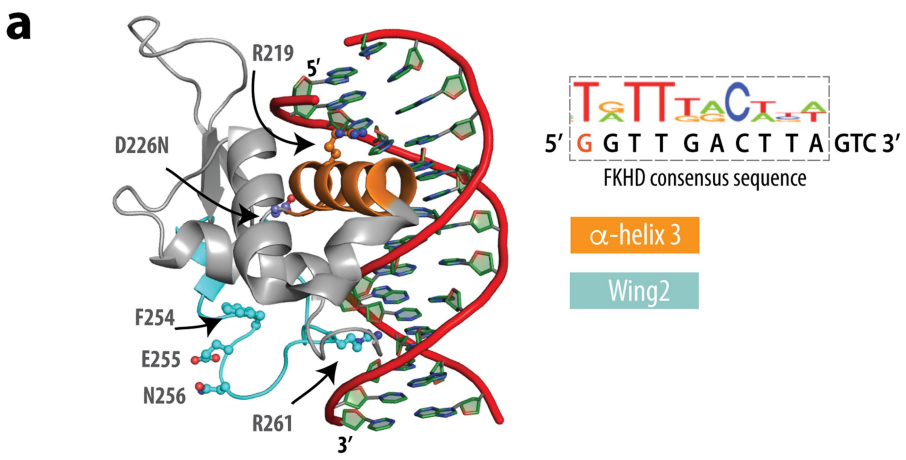
Author contributions E.J.A. and C.L.S. conceived and oversaw the project, performed data interpretation, and co-wrote the manuscript. E.J.A. and E.H. performed immunoblots, in vitro cell growth assays, lumen formation assays, lumen area quantification, processed organoids for immunohistochemistry and prepared experiments for RNA-seq and ATAC-seq. E.J.A., E.H. and W.R.K. made 3D organoid lines. E.J.A., W.R.K., E.H. and P.A.W. cloned plasmid reagents. E.J.A., E.H., W.R.K. and Z.Z. carried out in vivo experiments. E.J.A., R.B. and D.L. performed RNA-seq analysis and GSEA. E.J.A., R.B., D.L., A.S., Y.L., E.D. and C.E.B. performed analysis of human prostate cancer cohorts. A.G. optimized and carried out ATAC and ChIP protocols. R.D., S.C., H.C. and C.S.L. carried out ATAC-seq and ChIP–seq data analysis. All authors made intellectual contributions and reviewed the manuscript.

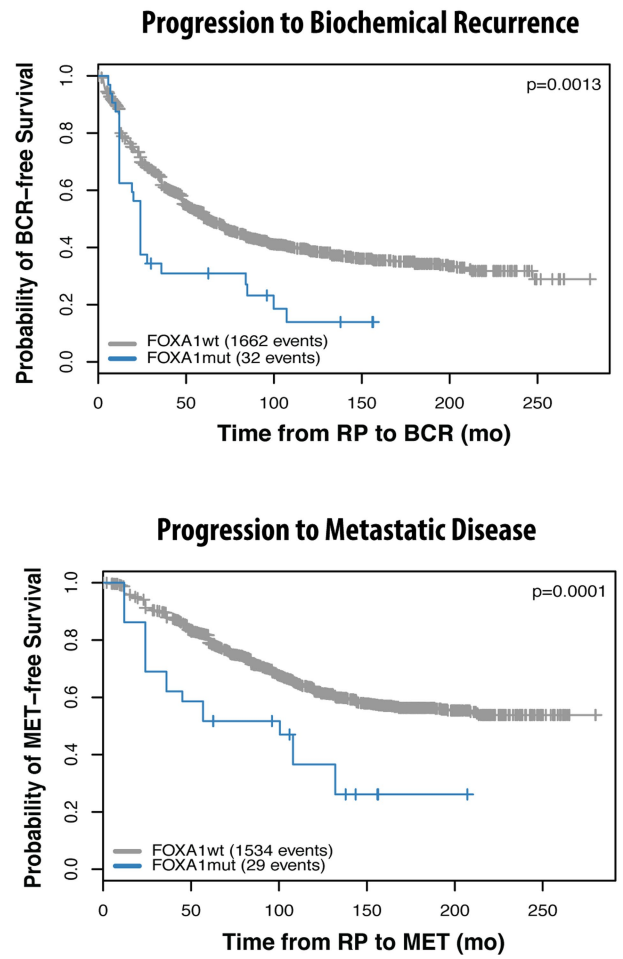
Competing interests C.L.S. serves on the board of directors of Novartis, is a co-founder of ORIC Pharm and co-inventor of enzalutamide and apalutamide. He is a science advisor to Agios, Beigene, Blueprint, Column Group, Foghorn, Housey Pharma, Nextech, KSQ, Petra and PMV. He was a co-founder of Seragon, purchased by Genentech/Roche in 2014. The other authors declare no competing interests.

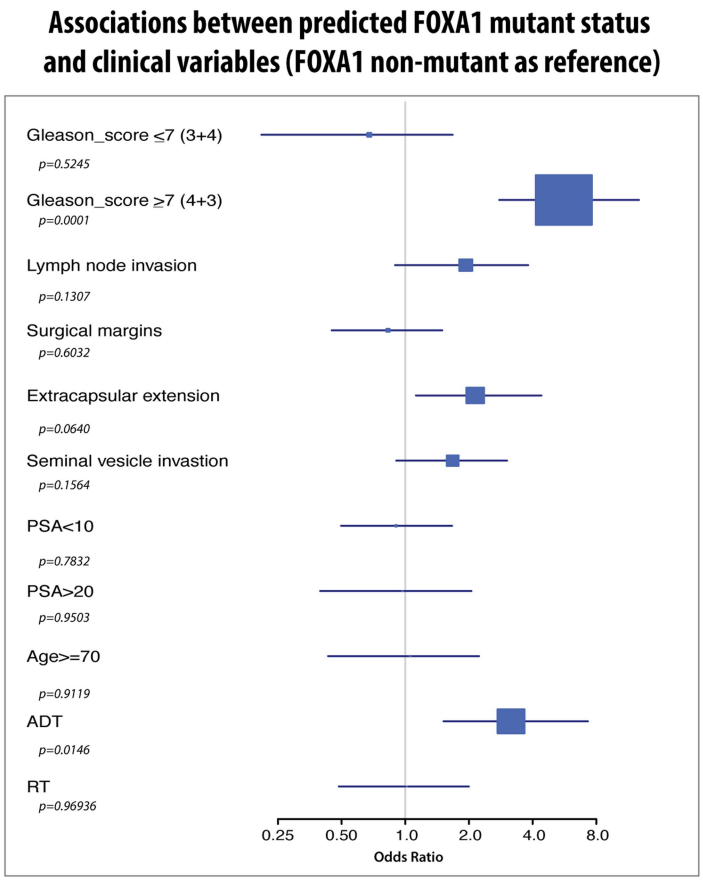
Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1318-9.

Correspondence and requests for materials should be addressed to C.L.S.

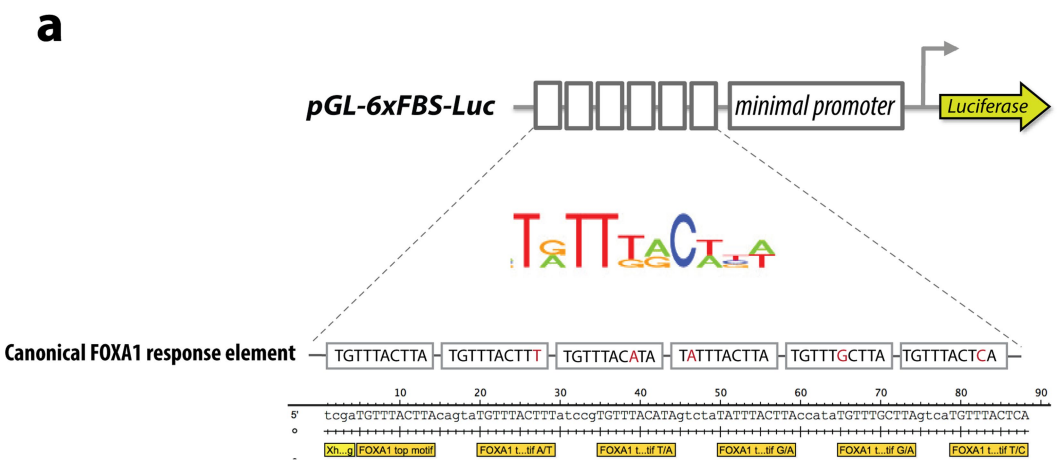


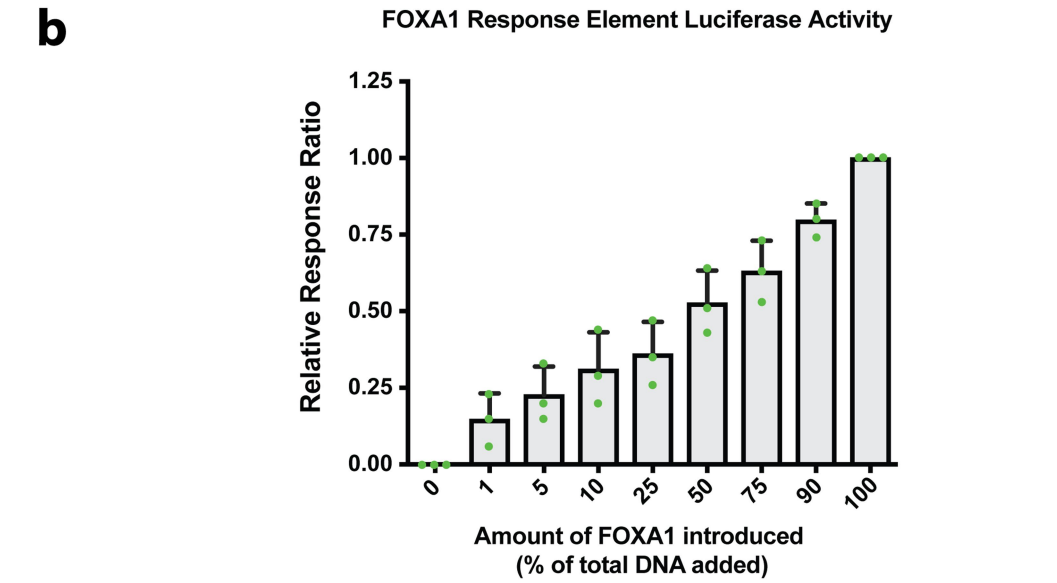


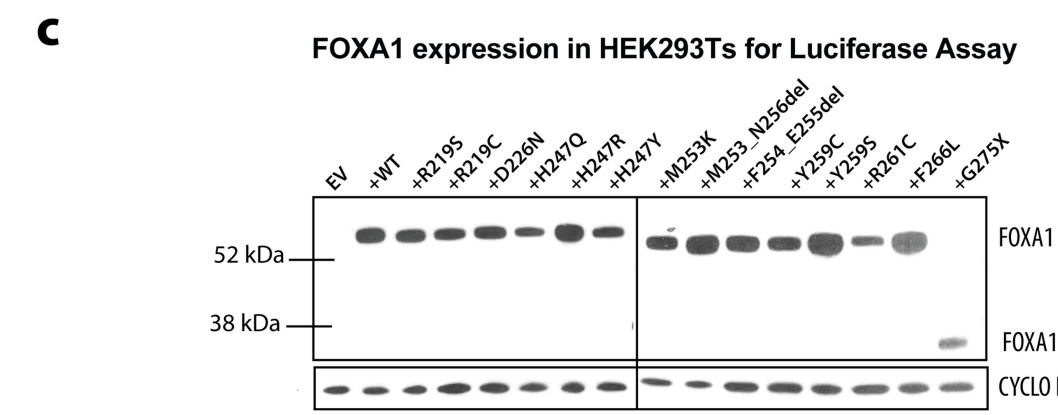


Extended Data Fig. 1 | See next page for caption.

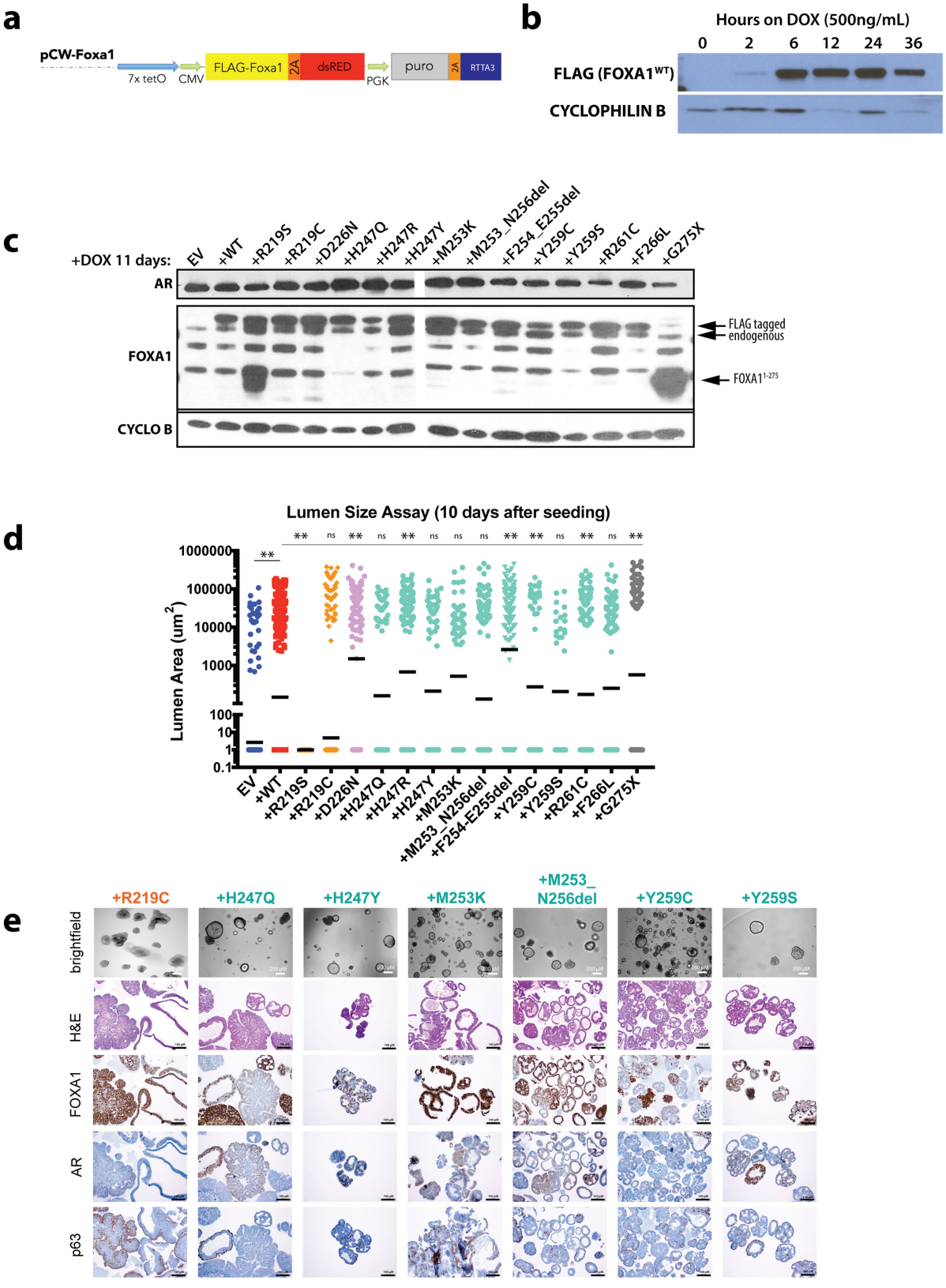
Extended Data Fig. 1 | Patients with predicted *FOXA1* mutant status have worse outcomes. a, Co-crystal structure of the FKHD domain of FOXA3 in complex with DNA resembling the FKHD consensus sequence (PDB 1VTN), with residues and folds of interest indicated, including α-helix 3 (orange), which sits in the major groove of DNA, and Wing2 (cyan), which undergoes frequent mutation in prostate cancer. **b**, Kaplan–Meier plot showing significantly different clinical outcomes of time to biochemical recurrence (BCR, top) or progression to metastatic disease (MET, bottom) for predicted *FOXA1* mutant cases vs wild type in the GRID cohort. The difference of MET and BCR survival curves was tested with the R survdiff function, using the G-rho family of tests, without adjustments for multiple comparisons. RP, radical prostatectomy. **c**, Associations between predicted *FOXA1* mutation status and clinical variables using univariate analysis of the GRID cohort, with *FOXA1* wild type as reference. The GRID cohort included 1,626 radical prostatectomy tumour samples. The centre values represent the median odds ratio via univariate analysis. The error bars represent first and third quartiles of odds ratio. The lines represent minimum and maximum odds ratio. Univariate logistic regression analyses were performed on the GRID cohort to test the statistical association between *FOXA1* mutant status and clinical variables via generalized linear test, without adjustments for multiple comparisons. The test was two-sided with the significance level of $P < 0.05$ as the cut-off. ADT, androgen deprivation therapy; PSA, prostate-specific antigen; RT, radiotherapy.





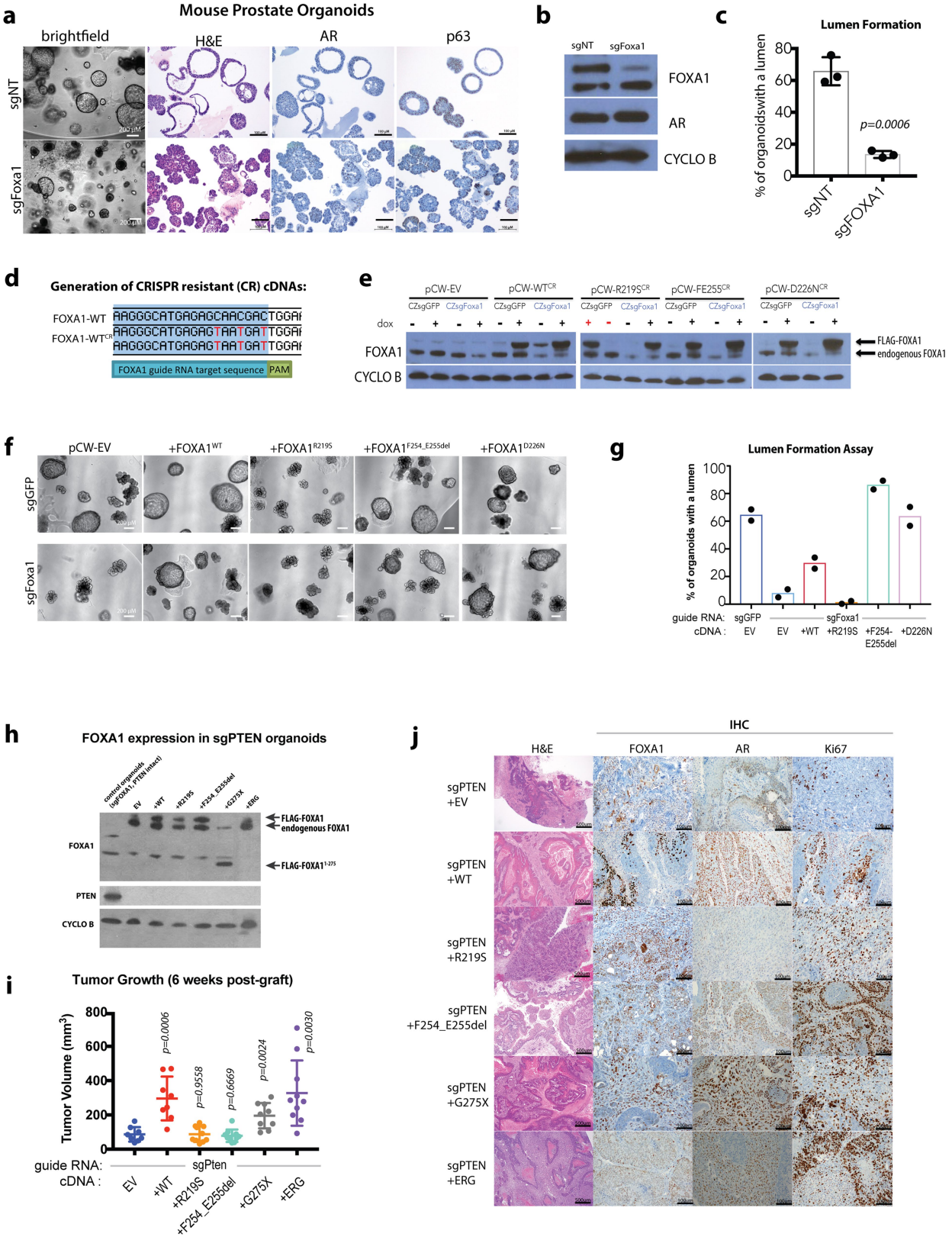


Extended Data Fig. 2 | Details of FOXA1 luciferase reporter assay. **a**, Schematic of FOXA1 luciferase reporter, depicting the modified response elements (at wobble positions within the canonical FOXA1 motif) cloned in tandem upstream of a minimal promoter driving luciferase expression. **b**, Dose–response curve of FOXA1 luciferase reporter activity in response to increased amounts of *Foxa1*^{WT} cDNA introduced into the system, expressed as a relative response ratio with 100% *Foxa1*^{WT} cDNA set to 1 and 0% *Foxa1*^{WT} cDNA (100% 'stuffer' DNA) set to 0. Data are from three biological replicates, central line and error bars represent mean ± standard deviation. **c**, Western blot of allelic series of *FOXA1* mutants in HEK293T cells 24 h after transfection with equal amounts of cDNA as used in FOXA1 luciferase reporter assay. The ΔF254/E255 and ΔM253–N256 mutants are shown as F254_E255del and M253_N256del, respectively, in the Extended Data. CYCLO B, loading control cyclophilin B. Representative blot, experiment repeated three independent times with similar results. For source gel data, see Supplementary Fig. 1.



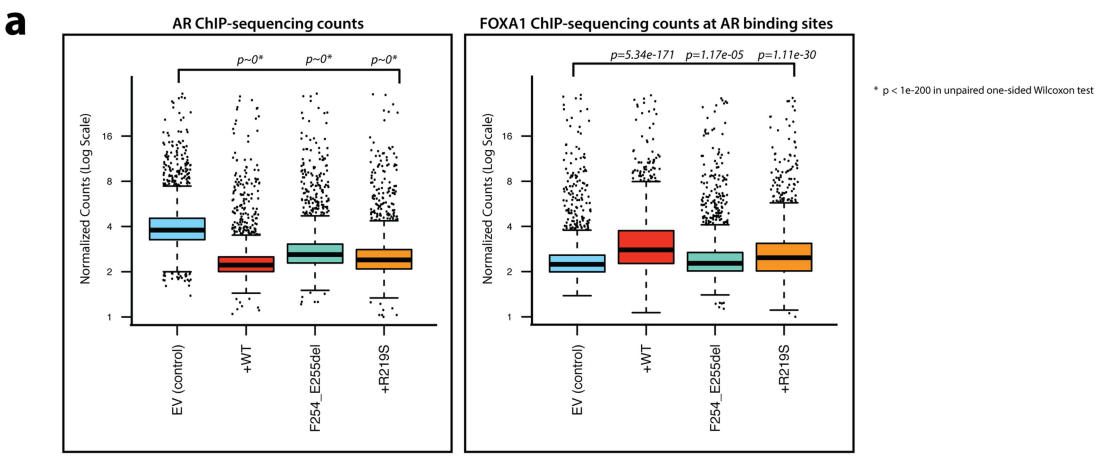
Extended Data Fig. 3 | See next page for caption.

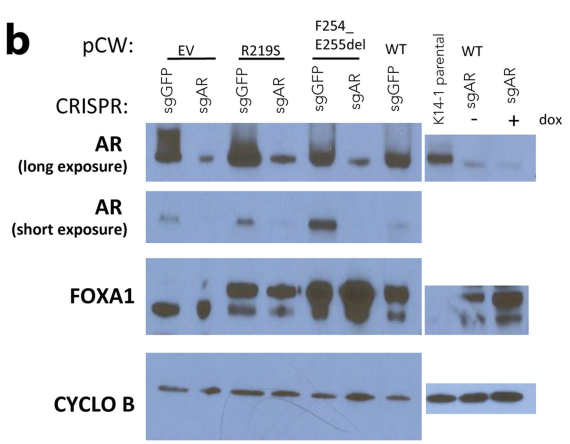
Extended Data Fig. 3 | Inducible overexpression of FOXA1 variants influences organoid lumen size and morphology. a, Schematic of doxycycline-inducible pCW-FOXA1 constructs used in the study. **b**, Western blot analysis of lysates from pCW-FOXA1(WT) organoids following acute doxycycline treatment. Representative blot, experiment repeated two independent times with similar results. For source gel data, see Supplementary Fig. 1. **c**, Western blot analysis of lysates from organoids following long-term doxycycline treatment. Sizes of endogenous and Flag-tagged FOXA1 are noted, as well as the smaller truncated form from G275X at the expected size ~38 kDa. Representative blot, experiment repeated three independent times with similar results. For source gel data, see Supplementary Fig. 1. **d**, Quantification of lumen areas measured at ten days after seeding. Solid black bar represents geometric mean. Values for sample size (indicated as dots) and *P* values are as follows: EV (292), +WT (284, $P < 0.0001$ over EV), +R219S (60, <0.0001), +F254_E255del (119, <0.0001), +D226N (120, <0.0001), +R261C (114, <0.0001), +R219C (333, 0.2915), +G275X (75, <0.0001), +M253_N256del (150, 0.2006), +M253K (63, 0.2343), +Y259S (32, 0.2045), +Y259C (45, 0.0082), +F266L (107, 0.1219), +H247Q (63, 0.8343), +H247R (180, <0.0001), +H247Y (71, 0.9104). All *P* values are relative to WT unless noted, calculated using unpaired, two-tailed Student's *t*-test. Colours represent location of mutation within FOXA1. **e**, Histology and immunohistochemistry of organoid lines overexpressing additional alleles of *FOXA1* via the doxycycline-inducible pCW vector 10 days after seeding. Images from a single biological experiment.

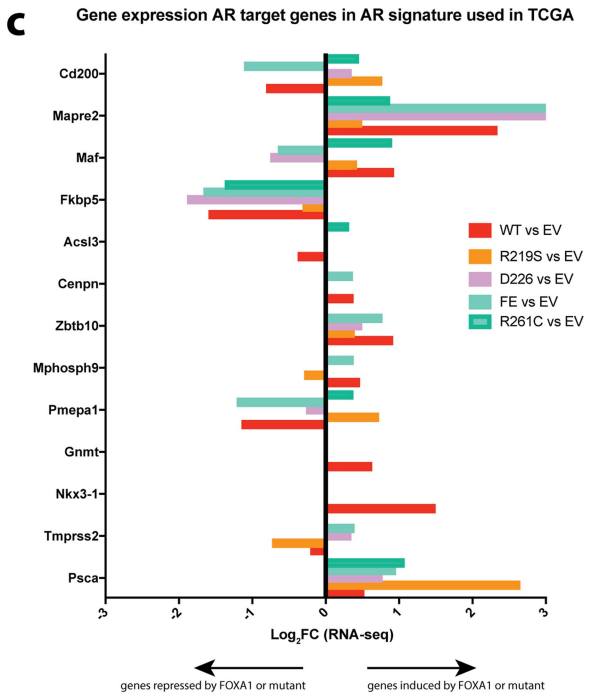


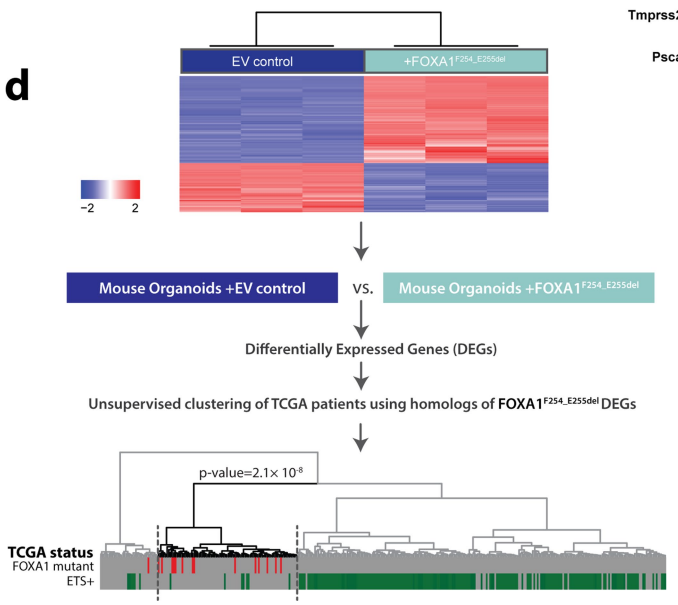
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Analysis of FOXA1 alterations in FOXA1-deleted or PTEN-deleted contexts. a, CRISPR–Cas9-mediated knockdown of *Foxa1* results in a markedly altered morphology. Organoids lacking *Foxa1* (sgFoxa1) have a reduced capacity to form lumens while maintaining expression of AR and the basal marker p63. sgNT (guide RNA targeting human gene AAVS1) serves as a negative control. **b**, Western blot analysis of lysates from organoids carrying control guide RNA (sgNT) or guide RNA targeting *Foxa1*. Representative blot, experiment repeated three times with similar results. For source gel data, see Supplementary Fig. 1. **c**, Quantification of organoids containing lumens, seven days after trypsinization in normal organoid media. Data are from three biological replicates, bars represent mean ± standard deviation, *P* value calculated using unpaired, two-tailed Student's *t*-test. **d**, Sequence indicating the location of three silent point mutations introduced upstream of the PAM sequence for *Foxa1*-targeting RNA sgFoxa1_1. **e**, Western blot analysis of lysates from organoids carrying either CRISPR-Zeo-sgGFP (CZsgGFP) or CRISPR-Zeo-sgFoxa1_1 (CZsgFoxa1) in addition to the pCW construct indicated, either EV or with a *Foxa1* allele present, plus or minus doxycycline treatment for ten days. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. **f**, Images of organoid lines carrying various combinations of guide RNA and cDNAs, ten days after doxycycline treatment. **g**, Quantification of lumen-containing organoids in lines with endogenous *Foxa1* deleted via CRISPR–Cas9 (sgFoxa1, sgGFP as control guide) and overexpression of CRIPSR-resistant $Foxa1^{WT}$ or mutant cDNA ten days after seeding. Data are from two biological replicates, bars represent mean. **h**, Western blot analysis of lysates from PTEN-deficient organoids grafted into mice, with doxycycline-induced overexpression of appropriate FOXA1 variants. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. **i**, Overexpression of FOXA1(WT) or FOXA1(G275X) in sgPTEN organoids promotes tumour growth in mice at six weeks after engraftment into the flank of NOD scid gamma mice. Data are from the following number of tumours: EV = 8, +WT = 8, +R219S = 10, +F254_E255del = 10, +G275X = 9, +ERG = 10. Data are mean ± s.d., *P* values calculated using unpaired, two-tailed Student's *t*-test vs EV. Colours represent location of mutation within FOXA1. **j**, Representative histology and immunohistochemistry (IHC) of a single tumour for given PTEN-deficient, FOXA1-expressing lines. Histology and immunohistochemistry were performed on 5–9 tumours per line, from a single in vivo experiment, with similar results.



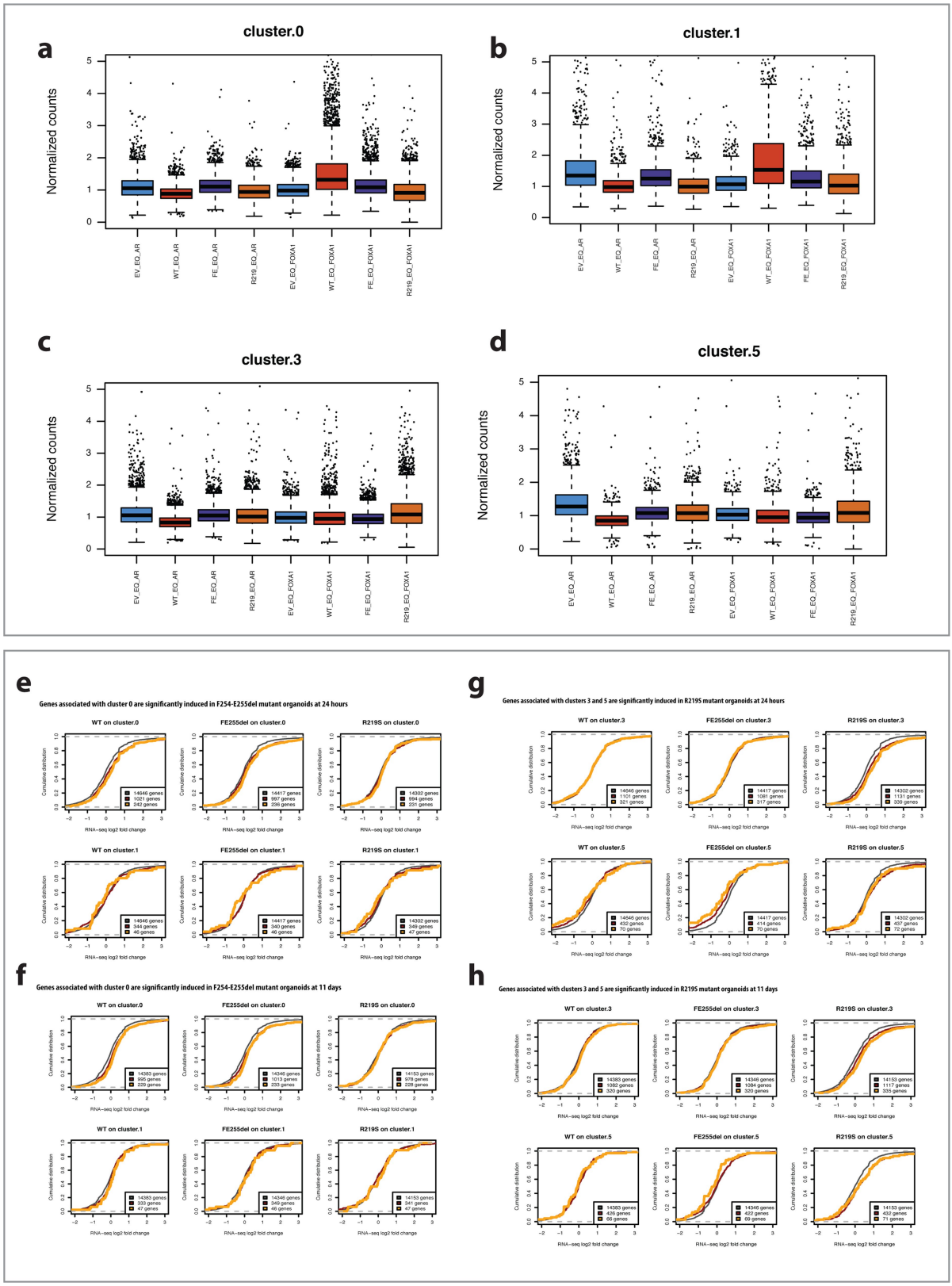






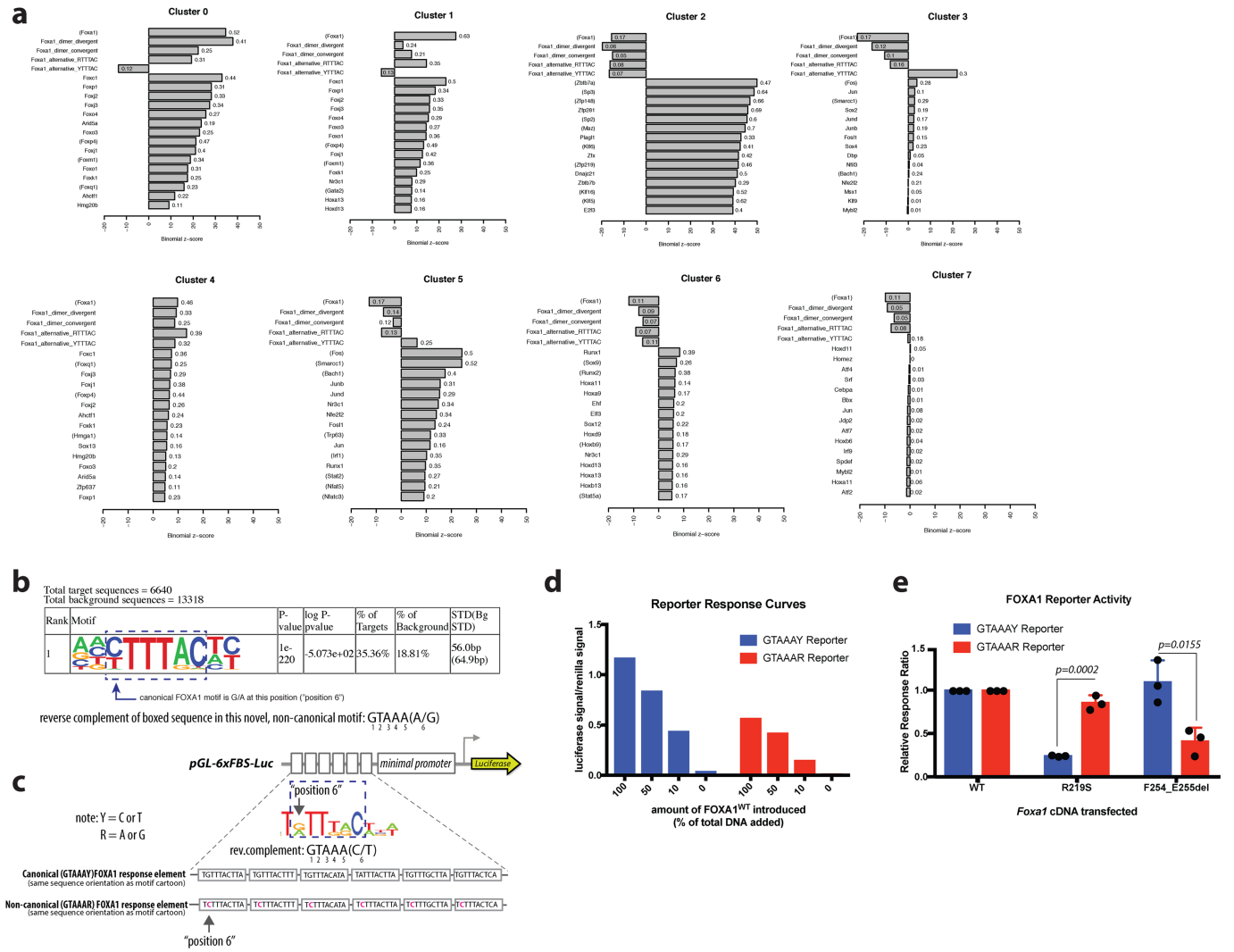
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Analysis of the interplay between AR and FOXA1 in mouse organoids expressing FOXA1 variants. a, Box plot representations of normalized counts from AR (left) and FOXA1 (right) ChIP–seq shown in Fig. 3a to quantify the reduction in AR binding following FOXA1 wild-type or mutant overexpression, and the increase in FOXA1 wild-type binding at those sites where AR is lost. Box: 25th to 75th percentile; band: median; top whisker: 75th percentile plus 1.5 times interquartile range; bottom whisker: 25th percentile minus 1.5 times interquartile range. Sample size = 2,914 peaks. *P* values calculated using an unpaired, one-sided Wilcoxon test. **b**, Western blot analysis of lysates from AR-deficient organoids generated using CRISPR–Cas9 carrying representative *Foxa1* alleles. Levels are significantly reduced but AR is not completely absent (as seen on the long exposure); this is a bulk population rather than single cell clones and thus a small number of cells escaped CRISPR–Cas9-mediated *Ar* deletion. Cells were treated with doxycycline for at least ten days. Representative blot, experiment repeated twice with similar results. For source gel data, see Supplementary Fig. 1. **c**, Expression of mouse orthologues of AR-target genes found in the AR signature used in TCGA cohort analysis based on mouse organoid RNA-seq. Genes depicted are those that have a mouse orthologue of the human gene found in the signature, and a significant expression change (DESeq2 adjusted $P < 0.05$) compared to EV control at 11 days of doxycycline treatment, as well as *Psca*, an AR-target gene expressed in mouse organoids. Data are from RNA-seq of three biological replicates. FE, F254_E255del. **d**, FOXA1(F254_E255del) signature can predict mutant tumours in TCGA. Hierarchical clustering and heat map of significantly differentially expressed genes between mouse FOXA1(F254_E255del) organoids and EV control (FDR $\leq 1 \times 10^{-10}$). Human homologues of differentially expressed genes (DEGs) from this analysis were used to cluster *FOXA1* mutant tumours ($n = 14$) and can detect nearly all *FOXA1* mutant human tumours ($P = 2.1 \times 10^{-8}$) out of the 333 TCGA samples, 199 of which are ETS^{+}. Two-sided Fisher's exact test was used to test the enrichment of *FOXA1* mutant samples within each sub-cluster, without adjustments for multiple comparisons.

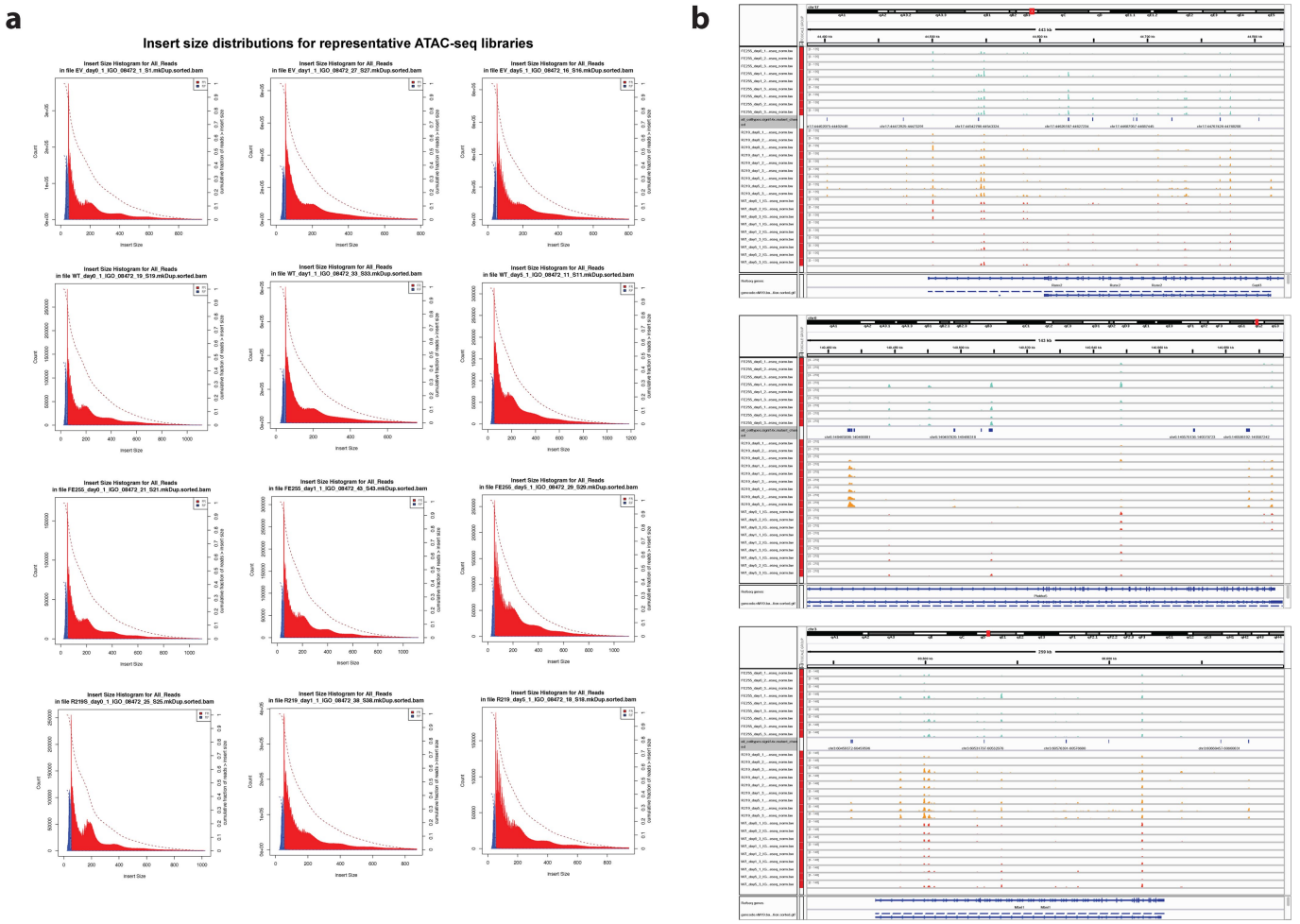


Extended Data Fig. 6 | See next page for caption.

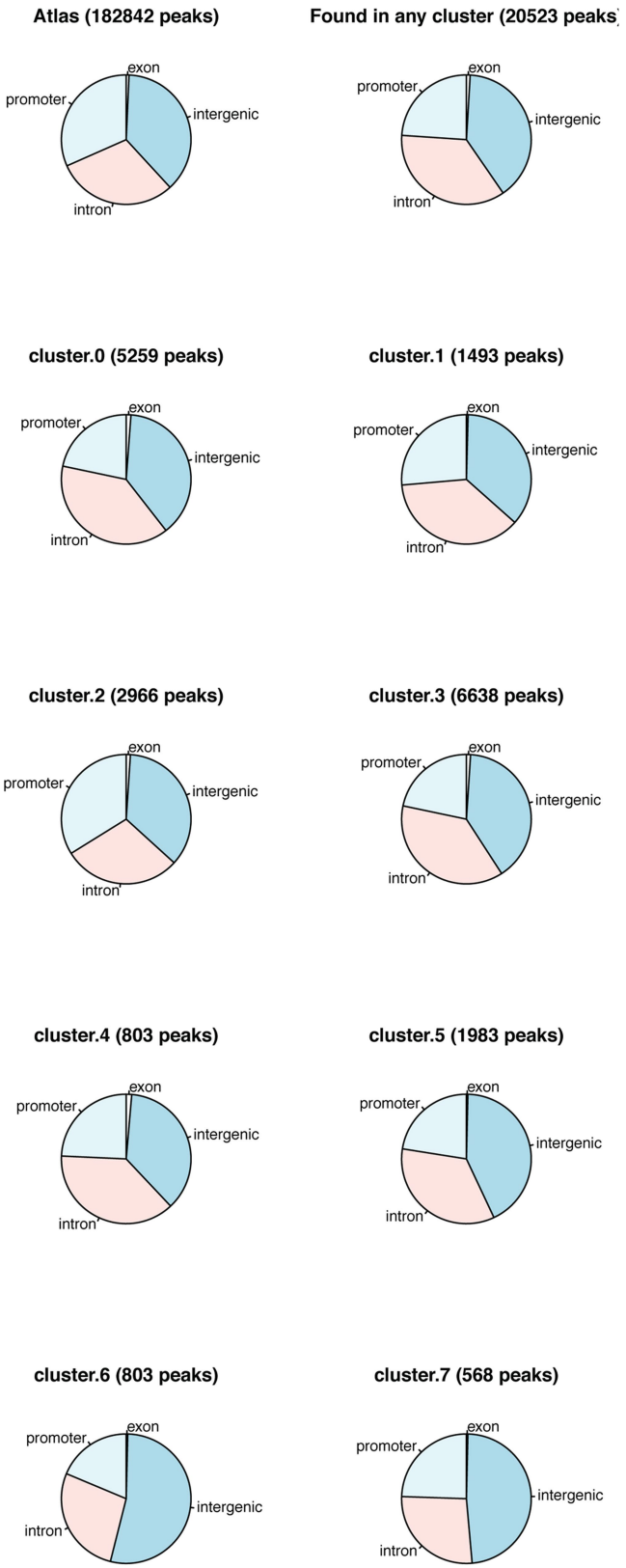
Extended Data Fig. 6 | Integrated analysis of ChIP–seq, ATAC-seq and RNA-seq data in *FOXA1* mutant organoid lines. a, Cluster 0 peaks have higher FOXA1 ChIP–seq signal in F254_E255del mutant organoid than empty vector control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 0, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher in F254_E255del (FE) and in WT compared to EV control (*P* values are listed in Supplementary Table 11). Sample size = 5,260 peaks. **b**, Cluster 1 peaks have higher FOXA1 ChIP–seq signal and lower AR ChIP–seq signal in FOXA1(WT)-overexpressing organoids than in EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 1, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher, and AR ChIP signal significantly lower, in WT compared to EV control. Sample size = 1,493 peaks. **c**, Cluster 3 peaks have higher FOXA1 ChIP–seq signal in R219S organoid than EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 3, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher in R219S compared to EV control. Sample size = 6,641 peaks. **d**, Cluster 5 peaks have higher FOXA1 ChIP–seq signal and lower AR ChIP–seq signal in R219S organoid than EV control. Box plots show normalized day five AR ChIP–seq signal and FOXA1 ChIP–seq signal across different organoid lines at peaks from cluster 5, where normalization is based on background ChIP signal. FOXA1 ChIP signal is significantly higher, and AR ChIP signal significantly lower, in R219S compared to EV control. Sample size = 1,983 peaks. In **a**–**d**, box: 25th to 75th percentile; band: median; top whisker: 75th percentile plus 1.5 times interquartile range; bottom whisker: 25th percentile minus 1.5 times interquartile range. *P* values calculated using an unpaired, one-sided Wilcoxon test. **e**, Genes associated with cluster 0 are significantly induced in F254_E255del mutant organoids. Top, plots show empirical cumulative distribution of $\log_2$ expression changes at 24 h vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 0 (cluster 0-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 0-associated genes, yellow). Cluster 0-associated genes show strong expression induction compared to all genes in F254_E255del as well as in WT (red vs black) but not in R219. Bottom, As a control, similar cumulative $\log_2$ expression changes for cluster 1-associated genes (red) or strong cluster 1-associated genes (yellow) do not show significant induction in F254_E255del. All *P* values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. **f**, Genes associated with cluster 0 are significantly induced in F254–E255del mutant organoids. Top, plots show empirical cumulative distribution of $\log_2$ expression changes at 11 days vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 0 (cluster 0-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 0-associated genes, yellow). Cluster 0-associated genes show strong expression induction compared to all genes in F254_E255del as well as in WT but not in R219. Bottom, As a control, similar cumulative $\log_2$ expression changes for cluster 1-associated genes (red) or strong cluster 1-associated genes (yellow) do not show significant induction in F254_E255del. All *P* values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. **g**, Genes associated with clusters 3 and 5 are significantly induced in R219S mutant organoid. Top, plots show empirical cumulative distribution of $\log_2$ expression changes at 24 h vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 3 (cluster 3-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 3-associated genes, yellow). Cluster 3-associated genes show strong expression induction compared to all genes in R219S but not in WT or F254_E255del. Bottom, similar analysis for cumulative $\log_2$ expression changes for cluster 5-associated genes (red) and strong cluster 5-associated genes (yellow). These genes are significantly induced in R219S and repressed in F254_E255del in WT for this time point. All *P* values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests. **h**, Genes associated with clusters 3 and 5 are significantly induced in R219S mutant organoid. Top, Plots show empirical cumulative distribution of $\log_2$ expression changes at day 11 vs day 0 in WT (left), F254_E255del mutant (middle) and R219S mutant (right) organoids for all expressed genes (black), genes associated with at least one ATAC-seq peak in cluster 3 (cluster 3-associated genes, red), and the top quartile of these genes based on number of assigned cluster 0 peaks (strong cluster 3-associated genes, yellow). Cluster 3-associated genes show strong expression induction compared to all genes in R219S but not in WT or F254_E255del. Bottom, similar analysis for cumulative $\log_2$ expression changes for cluster 5-associated genes (red) and strong cluster 5-associated genes (yellow). These genes are significantly induced in R219S and repressed in F254_E255del. All *P* values are listed in Supplementary Table 12 and are one-sided Wilcoxon rank-sum tests.



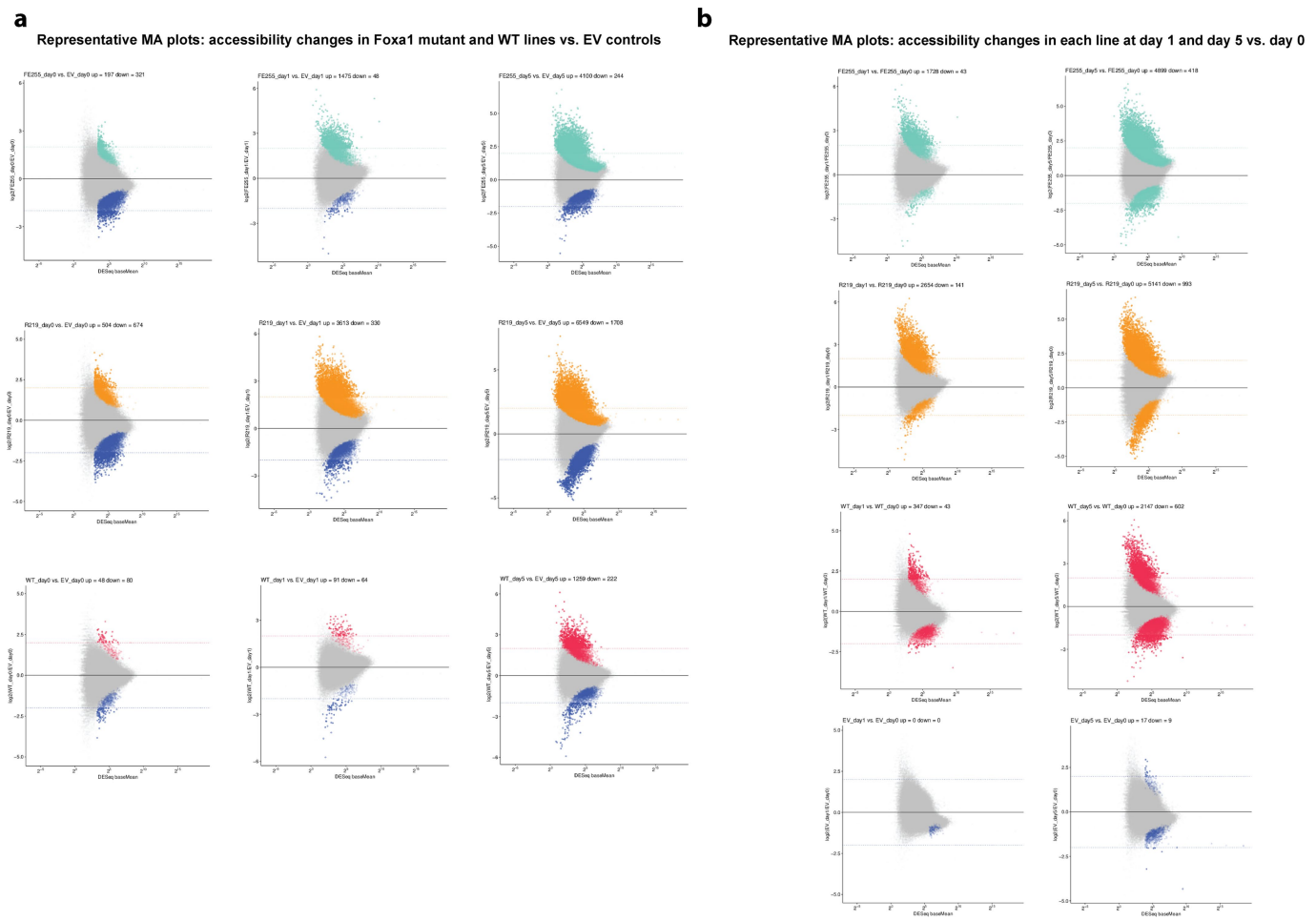
Extended Data Fig. 7 | Motif analysis of ATAC-sequencing and modification of FOXA1 reporter assay for evaluation of non-canonical FOXA1 motif. a, FIMO motif analysis of ATAC-seq clusters. Summary of motif enrichments and depletion results for each cluster relative to the background of all differentially accessible peaks, as reported by binomial *Z*-score. The top 15 enriched database motifs for expressed transcription factors are shown for each cluster. In addition, enrichment–depletion results for four additional FOXA1-related motifs are shown: convergent and divergent dimer motifs, and altered FOXA1 core binding motifs with either G/A or C/T at position 6. Transcription factors in parentheses represent motifs inferred from other species. Complete lists can be found in Supplementary Tables 3–10. **b**, Top motif identified de novo using HOMER on ATAC-seq cluster 3 (R219S-specific) with motif core indicated, and variation from canonical FOXA1 motif depicted. *P* values derived from one-sided binomial test. **c**, Schematic of reporter design. The canonical response element reporter is the same reporter used in Fig. 2, with various iterations of the canonical FOXA1 motif in tandem. The non-canonical motif has substitutions at position 6, indicated in pink, to reflect the newly identified motif enriched in cluster 3 of ATAC-seq. Note that the orientation of the upper motif cartoon and the sequence in the reporter schematic are the reverse complement of the motif identified by HOMER (GTAAAR). Modified base is noted in position 6. **d**, Dose–response curve for activity of both FOXA1 luciferase reporters in response to increased amounts of *Foxa1*^{WT} cDNA introduced into the system. Data shown are one representative biological replicate of three carried out, all showing same trends, but absolute luciferase/*Renilla* ratios vary from experiment to experiment. **e**, Results of reporter assays expressed as a relative response ratio, normalized to level of FOXA1(WT) activity for a given reporter. Data are from three biological replicates, mean ± s.d. Unpaired, two-tailed Student's *t*-test.



Extended Data Fig. 8 | Insert size distributions for ATAC-seq experiments, and track figures demonstrating peak reproducibility across ATAC-seq replicates. a, Representative insert size distributions computed from individual ATAC-seq experiments based on aligned read pairs, showing modes corresponding to nucleosome-free regions, mono-nucleosomal fragments, and di-nucleosomal fragments. **b**, Signal tracks for individual replicate ATAC-seq experiments at the *Runx2*, *Plekha5* and *Mbnl1* loci show reproducibility of accessibility events. DEseq scaling factors estimated from the atlas of IDR-reproducible peaks were used for library size normalization.



Extended Data Fig. 9 | ATAC-seq peak annotation distributions. Fraction of peaks annotated as promoter, intergenic, intronic and exonic for full atlas of reproducible peaks, differentially accessible peaks, and by ATAC-seq cluster. See Supplementary Table 15 for full annotation counts.



Extended Data Fig. 10 | MA plots for differential accessibility analysis. **a**, MA plots for differential accessibility analysis relative to EV controls. Representative MA plots (log (fold change) vs mean read counts) for differential peak accessibility analysis of FOXA1 mutant- and WT-expressing organoid lines vs empty vector controls at day 0, day 1, and day 5. Peaks that are significantly differential at FDR-corrected $P < 0.05$ are shown in colour. Dotted lines at log (fold change) = 2 and log (fold change) = −2 show cut-offs used for requiring robust accessibility changes in pairwise comparisons. **b**, MA plots for differential accessibility analysis at different time points relative to day 0. Representative MA plots (log (fold change) vs mean read counts) for differential peak accessibility analysis in each organoid line at day 1 vs day 0 and day 5 vs day 0. In **a**, **b**, all sample sizes are $n = 183{,}093$ (number of peaks in the atlas). Peaks that are significantly differential at FDR-corrected $P < 0.05$ are shown in colour, using two-sided Wald test with Benjamini–Hochberg correction.

nature research

Corresponding author(s): Charles L. Sawyers

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	See methods "RNA isolation and sequencing", "Assay for Transposase Accessible chromatin (ATAC) coupled with Next Generation Sequencing (NGS)" and "Chromatin Immuno-Precipitation (ChIP) coupled with Next Generation Sequencing (NGS)" Specific software includes Nikon NIS elements software.
Data analysis	See methods, "ATAC data and preprocessing", "ATAC-seq atlas creation", "Assignment of ATAC-seq peaks to genes", "SCaPT development based on FOXA1 mutant transcriptional signature and SVM model" and "Prostate cancer molecular subclass prediction by decision tree" and "Bio-informatics analysis ChIP-seq." Specific software includes: STARaligner (v2.4.2a), featureCounts (v1.4.3), Picard (v1.83), R v3.4.0, Trim Galore! v0.4.5, CutAdapt v1.16, FastQC v0.11.7, Bowtie2 v2.3.4.1, DESeq2 v1.18.1, deeptools 3.0.2, Homer v4.10, Trimmomatic v0.35, IDR (v2.0.2).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data will be made available from the authors upon reasonable request. The described RNA-seq and ChIP–seq and ATAC-seq data will be deposited in the Gene Expression Omnibus in the future.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample size calculation was done, for most assays, 3 biological replicates were carried out. For organoid lines, there are often more than 3 because lines were often included in a replicate experiment before results of previous experiment was analyzed, resulting in a greater number of replicates than what was likely needed.
Data exclusions	Entire growth assays were excluded if the control line did not grow as expected (i.e. didn't recover well from sorting so there was no growth seen). In that situation, data from entire experiment (all lines) were excluded. Mouse tumor measurements were excluded from data if it was noted at the time of injection that there was a technical issue. Both of these circumstances were pre-established and were held true for all lines.
Replication	All findings were reproducible, as more than 1 biological replicate was generally carried out, or the sample sizes were large enough to ensure good representation of the population behavior.
Randomization	No randomization was done for our experiments because there were no drug treatment groups, and all work was done in vitro.
Blinding	For lumen size assays, actual measurements were carried out by a member of the lab who did not have any knowledge of which alleles were associated with growth/lumen formation phenotypes. All other experiments were not blinded and it was not necessary to be as they were much more quantitative and less subjective.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☐	☒ Animals and other organisms
☒	☐ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	All unique materials are readily available upon request.

Antibodies

Antibodies used	AR (1:1,000, N-20, Santa Cruz), FOXA1 (1:1000, Ab2, Sigma), Cyclophilin B (1:1000, EPR12703(B), Abcam), or FLAG (1:1000, M2, Sigma), HNF-3 alpha/FoxA1 Antibody (3B3NB) 5ug/mL (Novus Biologicals), p63 (1:800, 4A4, Ventana), PTEN (1:1000, D4.3, Cell Signaling), Ki67 (Abcam #ab15580 at 1ug/ml). AR (1:1,000, ER179(2), Abcam)
Validation	AR (1:1,000, N-20, Santa Cruz, recommended for detection of AR of mouse, rat and human origin by WB, IP, IF, IHC(P) and ELISA) FOXA1 (1:1000, Ab2, Sigma, species reactivity: rat, human, mouse, canine; application: western blot, suitable) Cyclophilin B (1:1000, EPR12703(B), Abcam, Tested applications: Suitable for: WB, Unsuitable for: Flow Cyt,ICC/IF or IHC-P Species reactivity: Reacts with: Mouse, Rat, Human) FLAG (1:1000, M2, Sigma, For highly sensitive and specific detection of FLAG fusion proteins by immunoblotting, immunoprecipitation, immunohistochemisty, immunofluorescence and immunocyotchemistry. Optimized for single banded detection of FLAG fusion proteins in mammalian, plant, and bacterial expression systems. HNF-3 alpha/FoxA1 Antibody (3B3NB) 5ug/mL (Novus Biologicals, Reactivity Hu, Mu, Rt; Applications WB, Flow, ICC/IF, IHC, IHC-P) p63 (1:800, 4A4, Ventana, Anti-p40 (BC28) Mouse Monoclonal Primary Antibody intended for laboratory use in the detection of the p40 protein in formalin-fixed, paraffin-embedded tissue. PTEN (1:1000, D4.3, Cell Signaling) Rabbit Monoclonal antibody with reactivity to mouse, human, rat, dog, monkey. Applications: WB, IP, IHC, IF, F, ChIP. Ki67 (Abcam #ab15580 at 1ug/ml, suitable for IHC - Wholemount, IHC-P, IHC-FrFl, Flow Cyt, ICC/IF, ICC)

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Primary mouse organoid lines were established in our laboratory for this study
Authentication	since these were developed in our lab, authentication was not necessary
Mycoplasma contamination	All organoid lines used were tested for mycoplasma contamination and were found to be negative
Commonly misidentified lines (See ICLAC register)	none of the cell lines used are in this register

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Mus Musculus strain NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ, males between 8-12 weeks at beginning of study (study duration ~12 weeks).
Wild animals	Study did not involve wild animals
Field-collected samples	Study did not involve field collected samples

ChIP-seq

Data deposition

☐ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☐ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	*For "Initial submission" or "Revised version" documents, provide reviewer access links. For your "Final submission" document, provide a link to the deposited data.*
Files in database submission	*Provide a list of all files available in the database submission.*
Genome browser session (e.g. UCSC)	*Provide a link to an anonymized genome browser session for "Initial submission" and "Revised version" documents only, to enable peer review. Write "no longer applicable" for "Final submission" documents.*

Methodology

Replicates	*Describe the experimental replicates, specifying number, type and replicate agreement.*
Sequencing depth	*Describe the sequencing depth for each experiment, providing the total number of reads, uniquely mapped reads, length of reads and whether they were paired- or single-end.*
Antibodies	HNF-3 alpha/FoxA1 Antibody (3B3NB) (Novus Biologicals), AR (ER179(2), Abcam)

Peak calling parameters	*Specify the command line program and parameters used for read mapping and peak calling, including the ChIP, control and index files used.*
Data quality	*Describe the methods used to ensure data quality in full detail, including how many peaks are at FDR 5% and above 5-fold enrichment.*
Software	*Describe the software used to collect and analyze the ChIP-seq data. For custom code that has been deposited into a community repository, provide accession details.*

LETTER

https://doi.org/10.1038/s41586-019-1347-4

Distinct structural classes of activating FOXA1 alterations in advanced prostate cancer

Abhijit Parolia[1,2,3,12], Marcin Cieslik[1,2,4,12], Shih-Chun Chu[1,2], Lanbo Xiao[1,2], Takahiro Ouchi[1,2], Yuping Zhang[1,2], Xiaoju Wang[1,2], Pankaj Vats[1,2], Xuhong Cao[1,2,5], Sethuramasundaram Pitchiaya[1,2], Fengyun Su[1,2], Rui Wang[1,2], Felix Y. Feng[6,7,8,9], Yi-Mi Wu[1,2], Robert J. Lonigro[1,2], Dan R. Robinson[1,2] & Arul M. Chinnaiyan[1,2,5,10,11]*

Forkhead box A1 (FOXA1) is a pioneer transcription factor that is essential for the normal development of several endoderm-derived organs, including the prostate gland[1,2]. FOXA1 is frequently mutated in hormone-receptor-driven prostate, breast, bladder and salivary-gland tumours[3–8]. However, it is unclear how FOXA1 alterations affect the development of cancer, and FOXA1 has previously been ascribed both tumour-suppressive[9–11] and oncogenic[12–14] roles. Here we assemble an aggregate cohort of 1,546 prostate cancers and show that FOXA1 alterations fall into three structural classes that diverge in clinical incidence and genetic co-alteration profiles, with a collective prevalence of 35%. Class-1 activating mutations originate in early prostate cancer without alterations in ETS or SPOP, selectively recur within the wing-2 region of the DNA-binding forkhead domain, enable enhanced chromatin mobility and binding frequency, and strongly transactivate a luminal androgen-receptor program of prostate oncogenesis. By contrast, class-2 activating mutations are acquired in metastatic prostate cancers, truncate the C-terminal domain of FOXA1, enable dominant chromatin binding by increasing DNA affinity and—through TLE3 inactivation—promote metastasis driven by the WNT pathway. Finally, class-3 genomic rearrangements are enriched in metastatic prostate cancers, consist of duplications and translocations within the *FOXA1* locus, and structurally reposition a conserved regulatory element—herein denoted FOXA1 mastermind (FOXMIND)—to drive overexpression of *FOXA1* or other oncogenes. Our study reaffirms the central role of FOXA1 in mediating oncogenesis driven by the androgen receptor, and provides mechanistic insights into how the classes of FOXA1 alteration promote the initiation and/or metastatic progression of prostate cancer. These results have direct implications for understanding the pathobiology of other hormone-receptor-driven cancers and rationalize the co-targeting of FOXA1 activity in therapeutic strategies.

FOXA1 independently binds to and de-compacts condensed chromatin to reveal the binding sites of partnering nuclear hormone receptors[15,16]. In prostate luminal epithelial cells, FOXA1 delimits tissue-specific enhancers[17], and reprograms androgen receptor (AR) activity in prostate cancer[14]. Accordingly, FOXA1 and AR are co-expressed in prostate cancer cells, in which FOXA1 activity is indispensable for cell survival and proliferation[14] (Extended Data Fig. 1a–i). It is notable that, in AR-dependent prostate cancer, *FOXA1* is the third most-highly mutated gene[4,5] and—as shown here—is located at one of the most-highly rearranged genomic loci. Counterintuitively, recent studies have suggested these alterations are inactivating[18,19] and have described FOXA1 as a tumour suppressor in AR-driven metastatic prostate cancer[9–11]. However, FOXA1 alterations have not yet been fully characterized or experimentally investigated in cancer.

To study these alterations, we first curated an aggregate cohort of prostate cancer that comprised 888 localized and 658 metastatic samples[4,5,8,20], of which 498 and 357, respectively, had matched RNA-sequencing (RNA-seq) data. Here, *FOXA1* mutations recurred at a frequency of 8–9% in primary disease, which increased to 12–13% in metastatic castration-resistant prostate cancer (mCRPC) (Fig. 1a, Extended Data Fig. 1j). RNA-seq calls of structural variants revealed a high prevalence (Fig. 1b, Supplementary Table 1) and density (Extended Data Fig. 1k) of rearrangements within the *FOXA1* locus. The presence of structural variants was confirmed by whole-exome and whole-genome sequencing (Extended Data Fig. 1l, m, Supplementary Tables 2, 3). Overall, we estimated the recurrence of *FOXA1* locus rearrangements to be 20–30% in mCRPC (Extended Data Fig. 1n). All *FOXA1* mutations were heterozygous and *FOXA1* itself was copy-amplified in over 50% of cases with no biallelic deletions (Extended Data Fig. 2a, b). We also found a stagewise increase in FOXA1 expression in prostate cancer (Extended Data Fig. 2c, Supplementary Discussion).

When we mapped mutations onto the protein domains of FOXA1, we found two structural patterns: (1) missense and in-frame insertion and deletion (indel) mutations were clustered at the C-terminal end of the forkhead domain (FKHD); and (2) truncating frameshift mutations were restricted to the C-terminal half of the protein (Fig. 1c). *FOXA1* structural variants predominantly consisted of tandem duplications and translocations, which clustered in close proximity to the *FOXA1* gene without disrupting its coding sequence (Fig. 1d). Thus, we categorized FOXA1 alterations into three structural classes: class 1, which comprises all the mutations within the FKHD; class 2, which comprises mutations in the C-terminal end after the FKHD; and class 3, which comprises structural variants within the *FOXA1* locus (Fig. 1c, d, Extended Data Fig. 2d). We also found similar classes of FOXA1 alterations in breast cancer (Extended Data Fig. 2e, f).

We found that the majority of *FOXA1* mutations in primary prostate cancer belonged to class 1, which showed no enrichment in the metastatic disease (Fig. 1e). Conversely, class-2 mutations were significantly enriched in metastatic prostate cancer; in the rare primary cases with class-2 mutations, the mutant allele was detected at sub-clonal frequencies (Fig. 1e, f, Extended Data Fig. 2g, h). We found no cases that possessed both class-1 and class-2 mutations. Class-3 structural variants were also significantly enriched in mCRPC (odds ratio = 3.46) (Fig. 1g). Overall, we found the cumulative frequency of FOXA1 alterations to be over 34% in mCRPC (Fig. 1h). Assessment of concurrent alterations revealed that class-1 mutations are mutually exclusive with other primary events (for example, ETS fusions) (odds ratio = 0.078), whereas class-2-mutant mCRPC are enriched for *RB1* deletions (odds ratio = 4.17) (Extended Data Fig. 2i, j). Both mutational classes were further enriched for alterations in DNA repair, mismatch repair and

[1]Michigan Center for Translational Pathology, University of Michigan, Ann Arbor, MI, USA. [2]Department of Pathology, University of Michigan, Ann Arbor, MI, USA. [3]Molecular and Cellular Pathology Program, University of Michigan, Ann Arbor, MI, USA. [4]Department of Computational Medicine and Bioinformatics, University of Michigan, Ann Arbor, MI, USA. [5]Howard Hughes Medical Institute, University of Michigan, Ann Arbor, MI, USA. [6]Helen Diller Family Comprehensive Cancer Center, University of California at San Francisco, San Francisco, CA, USA. [7]Department of Radiation Oncology, University of California at San Francisco, San Francisco, CA, USA. [8]Department of Urology, University of California at San Francisco, San Francisco, CA, USA. [9]Department of Medicine, University of California at San Francisco, San Francisco, CA, USA. [10]Department of Urology, University of Michigan, Ann Arbor, MI, USA. [11]Rogel Cancer Center, University of Michigan, Ann Arbor, MI, USA. [12]These authors contributed equally: Abhijit Parolia, Marcin Cieslik. *e-mail: arul@umich.edu

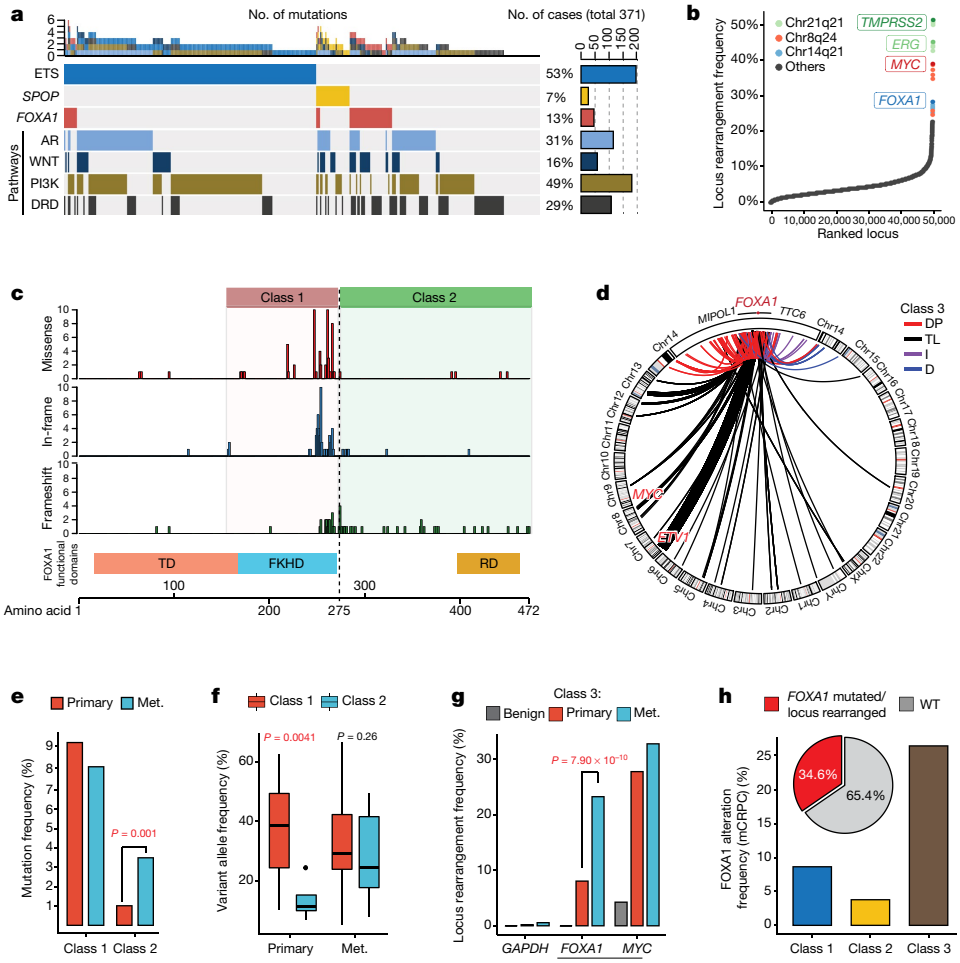


Fig. 1 | Structural classes of FOXA1 alterations. a, *FOXA1* mutations and key alterations in mCRPC. Alterations in ETS, AR, WNT, PI3K and DNA repair (DRD) were aggregated at the pathway or group level. **b**, Locus-level recurrence of RNA-seq structural variations. **c**, Structural classification of *FOXA1* mutations. TD, transactivation domain; RD, regulatory domain. **d**, Structural classification of *FOXA1* locus rearrangements. DP, tandem duplications; TL, translocations; I, inversions; D, deletions. **e**, Frequency of *FOXA1* mutational classes by prostate cancer stage ($n = 888$ primary, 658 metastatic (met.)) (two-sided Fisher's exact test). **f**, Variant allele frequency by stage and class (two-sided *t*-test). Box plot centre, median; box, quartiles 1–3, whiskers, quartiles 1–3 ± 1.5 × interquartile range (IQR). **g**, Locus-level recurrence of structural variants based on RNA-seq by prostate cancer stage (two-sided Fisher's exact test). **h**, Integrated (RNA-seq and whole-exome sequencing) recurrence of FOXA1-alteration classes in mCRPC (Stand Up 2 Cancer and Michigan Center for Translational Pathology (MCTP) cohort, $n = 370$).

WNT signalling pathways (Extended Data Fig. 2i, k), and had higher levels of expression of *FOXA1* mRNA relative to the wild-type cases (Extended Data Fig. 2l). Together, these data suggest that class-1 mutations emerge in localized prostate cancer, whereas class-2 and class-3 mutations are acquired or enriched, respectively, in the course of disease progression.

Class-1 mutations consist of missense and in-frame indels that cluster at the C-terminal edge of the winged-helix DNA-binding FKHD. The majority of the class-1 mutations were located either within the wing-2 region (residues 247–269) or a 3D hotspot that spatially protrudes towards wing 2[21] (Fig. 2a, b, Extended Data Fig. 3a, b). Notably, these mutations did not alter FKHD residues that make base-specific interactions with the DNA[22,23] (Fig. 2a, Extended Data Fig. 3c). In FOXA proteins, wing-2 residues make base-independent (that is, non-specific) contacts with the DNA backbone[23,24] that reportedly impede its nuclear movement[24]. Thus, we hypothesized that class-1 mutants with altered wing-2 regions would display faster nuclear mobility.

We cloned representative class-1 mutants of FOXA1: I176M (mutation of the 3D hotspot), R261G (missense) and R265–Q271del (in-frame deletion), all of which retained nuclear localization (Extended Data Fig. 3d). In fluorescence recovery after photobleaching (FRAP) assays, we found class-1 mutants had 5–6× faster nuclear mobility irrespective of the mutation type (Fig. 2c, d, Extended Data Fig. 3e, g). By contrast, class-2 mutants with intact wing-2 regions were sluggish in their nuclear movement (Fig. 2d, Extended Data Fig. 3f, g). Using single particle tracking, we verified that class-1 mutants have a higher overall rate of nuclear diffusion, with 3–4-fold fewer slow particles and shorter chromatin dwell times (Extended Data Fig. 3h, i). In chromatin immunoprecipitation with parallel DNA sequencing (ChIP–seq) assays, we found that ectopically expressed class-1 mutants in HEK293 cells bind DNA at the consensus FOXA1 motif (Extended Data Fig. 3j, k). In prostate cancer cells, the class-1 cistrome entirely overlapped with wild-type binding sites, with similar enrichment for FOXA1 and AR cofactor motifs, AR-binding sites and genomic distribution (Extended Data Fig. 3l–s). Furthermore, in growth rescue experiments using untranslated-region-specific small interfering (si) RNAs that targeted the endogenous *FOXA1* transcript, we found that exogenous class-1 mutants fully compensated for the wild-type protein (Extended Data Fig. 4a).

Next, we asked how class-1 mutations affect AR signalling. Similar to wild-type FOXA1, both class-1 and class-2 mutants interacted with the AR signalling complex (Extended Data Fig. 4b–d). In reporter assays, class-1 mutants induced 3–6-fold higher activation of AR signalling (Fig. 2e), which was evident even under stimulation with castrate levels of androgen or treatment with enzalutamide (Extended Data Fig. 4e, f). In parallel assays, class-2 mutants showed no differences relative to wild-type FOXA1 (Fig. 2e). Transcriptomic analyses of class-1 tumours

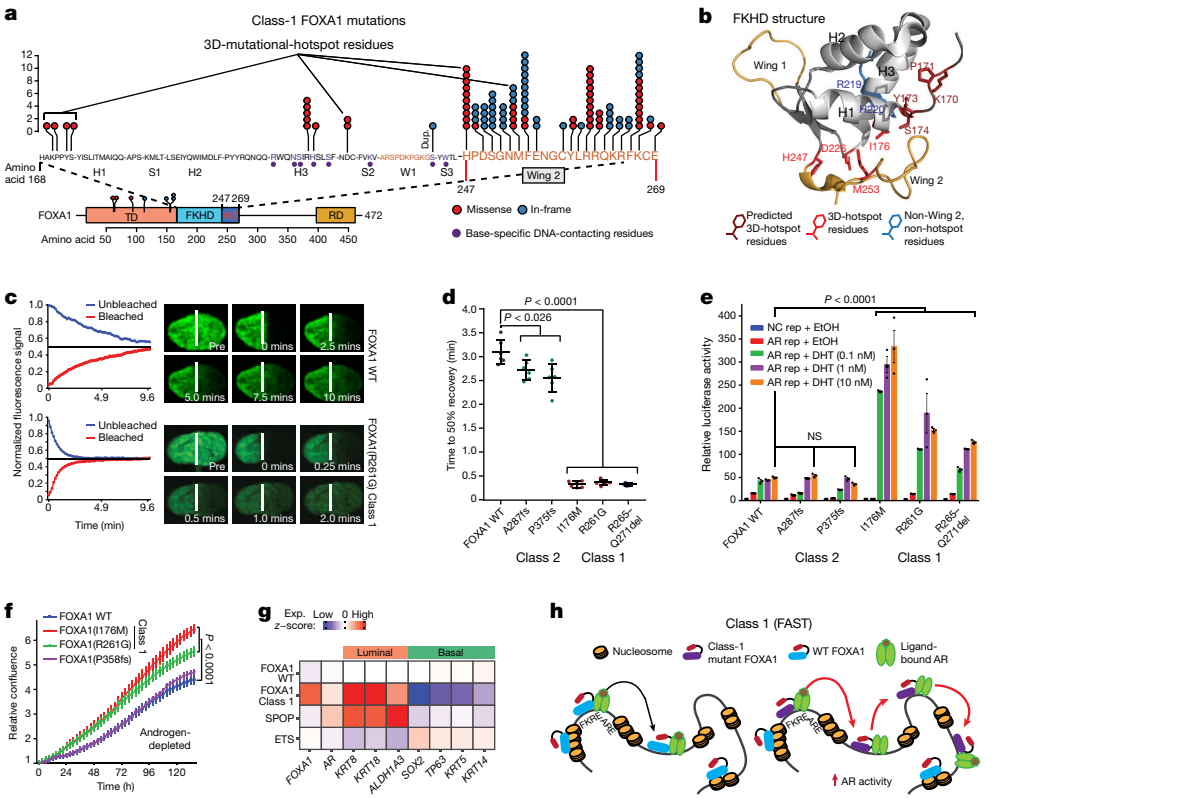


Fig. 2 | Functional characterization of class-1 mutations of *FOXA1*. **a**, Distribution of class-1 mutations on the protein map of FOXA1 functional domains and FKHD secondary structures. Dup., duplication. **b**, Crystal structure of the FKHD with visualization of non-wing-2 (that is, outside of amino acids 247–269) mutations. Mutations in the 3D hotspot are in red. **c**, FRAP kinetic plots (left) and representative time-lapse images from pre-bleaching to the equilibrated state (right; $n = 6$ biological replicates). Images are uniformly brightened for signal visualization. WT, wild type. **d**, FRAP durations until 50% recovery ($n = 6$ nuclei per variant). **e**, Negative control (NC) or AR reporter (rep) activity with overexpression of FOXA1 variants and dihydrotestosterone (DHT) stimulation ($n = 3$ biological replicates). **f**, Growth (IncuCyte) of 22RV1 cells that overexpress FOXA1 variants in androgen-depleted medium ($n = 5$ biological replicates). In **d**–**f**, mean ± s.e.m. is shown, and *P* values are from two-way analysis of variance (ANOVA) and Tukey's test. **g**, Relative expression of luminal and basal markers in class-1 ($n = 38$) tumours compared with wild-type ($n = 457$), SPOP ($n = 48$) and ETS ($n = 243$) primary prostate cancer tumours. **h**, Class-1 model. Wing-2-disrupted FOXA1 shows increased chromatin mobility and chromatin sampling frequency, which results in stronger transcriptional activation of oncogenic AR signalling. FKRE, forkhead-responsive element; ARE, androgen-responsive element.

from patients revealed the activation of hyperproliferative and pro-tumorigenesis pathways, and further enrichment of primary prostate cancer genes (Extended Data Fig. 4g–i). Notably, AR was predicted[25] to be the driver transcription factor for class-1 upregulated genes, which we experimentally confirmed for several targets (Extended Data Fig. 4j–l). Concordantly, overexpression of class-1 mutants in 22RV1 cells increased growth in androgen-depleted medium (Fig. 2f) but not in androgen-supplemented medium, and rescued proliferation upon treatment with enzalutamide (Extended Data Fig. 4m, n). For class-1 downregulated genes, the basal transcription factors TP63 and SOX2 were predicted to be transcriptional drivers (Extended Data Fig. 4j). Consistently, in class-1 specimens from patients, both of these transcription factors were significantly downregulated, with a concomitant downregulation of basal, and upregulation of luminal, markers (Fig. 2g, Extended Data Fig. 4o, p). In addition, class-1 tumours had a higher AR transcriptional signature, and a lower neuroendocrine transcriptional signature (Extended Data Fig. 4q). Together, these data suggest that class-1 mutations that alter the wing-2 region increase the nuclear speed and genome-scanning efficiency of FOXA1 without affecting its DNA sequence specificity (Supplementary Discussion), and drive a luminal AR program of prostate oncogenesis (Fig. 2h).

Class-2 mutations consist of frameshifting alterations that truncate the C-terminal regulatory domain of FOXA1 (Fig. 3a). Thus, we characterized the class-2 cistrome by using N-terminal and C-terminal antibodies, with the C-terminal antibody binding exclusively to wild-type FOXA1 (Extended Data Fig. 5a, b). Notably, mCRPC-derived LAPC4 cells endogenously contained a *FOXA1* class-2 mutation (that is, a frameshift at amino acid P358 (P358fs)), and both wild-type and mutant variants interacted with the AR complex (Extended Data Fig. 5c–f). However, in ChIP–seq assays, only the N-terminal antibody detected FOXA1 binding to the DNA. By contrast, N-terminal and C-terminal FOXA1 cistromes substantially overlapped in wild-type prostate cancer cells (Fig. 3b, Extended Data Fig. 5g–i). Even with 13-fold overexpression of wild-type FOXA1 in LAPC4 cells, the endogenous class-2 mutant retained its binding dominance (Fig. 3b, Extended Data Fig. 5j, k). Conversely, overexpression of the FOXA1(P358fs) mutant in LNCaP cells markedly diminished the endogenous wild-type cistrome (Fig. 3b). In in vitro assays, class-2 mutants showed markedly stronger binding to the *KLK3* enhancer element (Fig. 3c, Extended Data Fig. 6a–d), and biolayer interferometry confirmed that the FOXA1(P358fs) mutant has an approximately fivefold-higher DNA-binding affinity (Extended Data Fig. 6e). In CRISPR-engineered class-2-mutant 22RV1 clones (Extended Data Fig. 6f, g), FOXA1 ChIP–seq data reaffirmed the cistromic dominance of class-2 mutants (Fig. 3d). Knockdown of either mutant FOXA1 or AR in 22RV1 or LNCaP class-2 CRISPR clones significantly attenuated proliferation (Fig. 3e, Extended Data Fig. 6h, i). Consistently, in rescue experiments, the FOXA1(P358fs) mutant fully compensated for the loss of wild-type FOXA1 (Extended Data Fig. 4a).

The class-2 cistrome was considerably larger than the wild-type cistrome (Extended Data Fig. 6j–l), and the acquired sites were enriched for the CTCF motif and distal regulatory regions (Extended Data Fig. 7a–e, Supplementary Discussion). In transcriptomic and motif analyses of the class-2 clones, LEF and TCF were predicted as

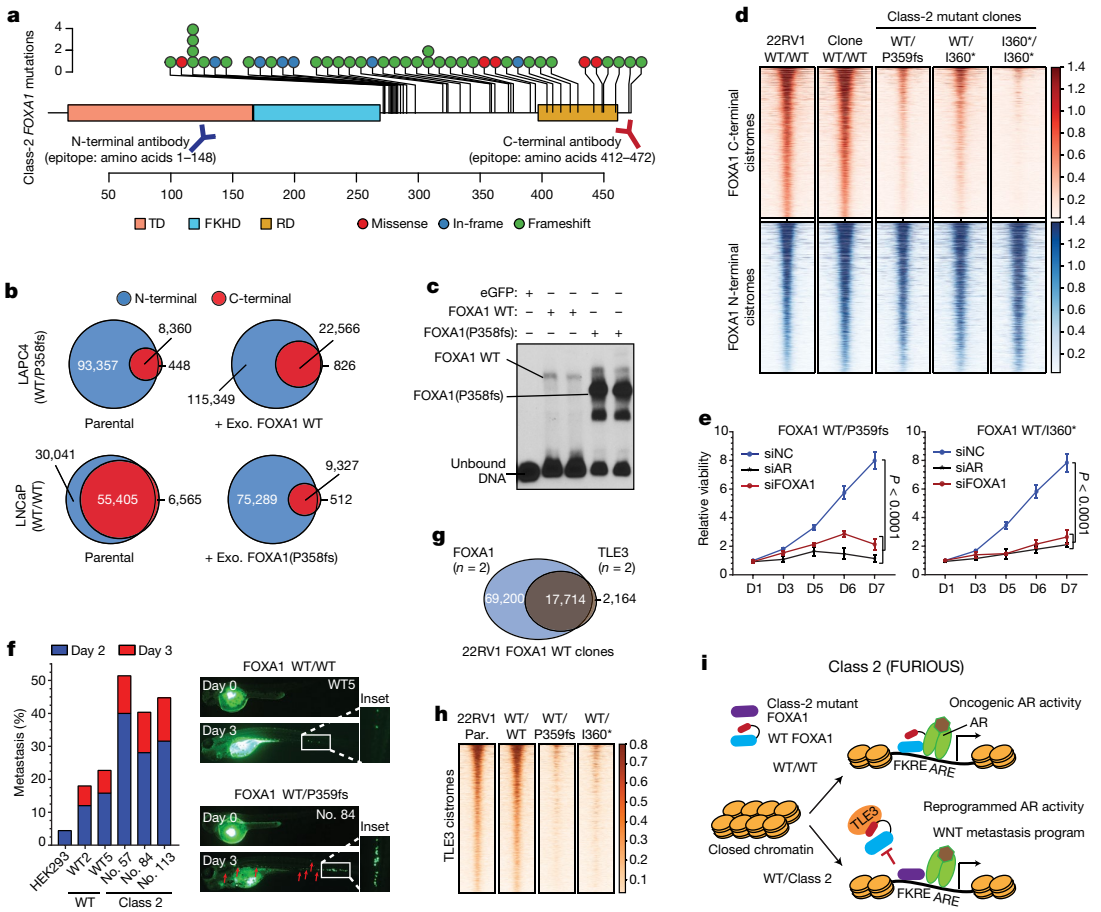


Fig. 3 | Functional characterization of class-2 mutations of *FOXA1*. **a**, Class-2 mutations and antibody epitopes on the protein map of FOXA1. **b**, N-terminal and C-terminal FOXA1 cistromes in FOXA1 wild-type (*FOXA1*$^{WT/WT}$ (WT/WT)) or mutant (*FOXA1*$^{WT/P358fs}$ (WT/P358fs)) prostate cancer cells that are untreated (left) or have exogenous (exo.) overexpression of FOXA1 variants (right). **c**, Electromobility shift of FOXA1 variants bound to the *KLK3* enhancer ($n = 3$ biological replicates). For gel source data, see Supplementary Fig. 1. **d**, FOXA1 ChIP–seq read-density heat maps in independent class-2-mutant 22RV1 CRISPR clones (*FOXA1*$^{WT/P359fs}$ (WT/P359fs), *FOXA1*$^{WT/I360*}$ (WT/I360*) and *FOXA1*^{I360*I360*} (I360*/I360*)). **e**, Growth of class-2-mutant 22RV1 clones treated with non-targeting (siNC), *AR*- or *FOXA1*-targeting siRNAs ($n = 5$ biological replicates; two-way ANOVA and Tukey's test). Mean ± s.e.m. is shown. D, day. **f**, Left, metastasis frequency in zebrafish embryos injected with HEK293 (negative control), wild-type 22RV1 clones or class-2-mutant 22RV1 clones ($n \geq 30$ embryos per group). Right, representative images of embryos, showing the disseminated prostate cancer cells. **g**, Overlap of wild-type FOXA1- and TLE3-binding sites in 22RV1 CRISPR clones ($n = 2$ biological replicates each). **h**, TLE3 ChIP–seq read-density heat maps in 22RV1 parental (par.) cells and distinct FOXA1 wild-type and class-2-mutant 22RV1 CRISPR clones. **i**, Class-2 model. Truncated FOXA1 shows dominant chromatin binding and displaces wild-type FOXA1 and TLE3 from the chromatin, which results in increased WNT signalling.

the top regulatory transcription factors for the upregulated genes (Extended Data Fig. 7g, h). The LEF–TCF complex is the primary nuclear effector of WNT signalling and remains inactive until it is bound by β-catenin[26]. Consistently, we found a marked accumulation of transcriptionally active β-catenin—that is, non-phosphorylated at S31, S37 and T41—in distinct mutant clones, as well as a concomitant increase in the expression of the WNT targets LEF1 and AXIN2 (Extended Data Fig. 7i, j). Class-2 clones showed 2–3-fold higher invasiveness in Boyden chamber assays (Extended Data Fig. 7k, l), and a higher rate and extent of metastatic dissemination in zebrafish embryos (Fig. 3f, Extended Data Fig. 7m). In these assays, class-1 mutant cells showed no differences relative to wild-type cells (Extended Data Fig. 7n). Furthermore, treatment with the WNT inhibitor XAV939 completely abrogated the class-2 invasive phenotype (Extended Data Fig. 7o). Investigating the mechanism that underlies this invasiveness, we found that FOXA1 transcriptionally activates and—through its C-terminal domain—recruits TLE3 (a bona fide WNT co-repressor[27]) to the chromatin (Extended Data Fig. 8a–e). Class-2 mutants had lost this interaction, which led to the untethering of TLE3 from chromatin and downstream activation of WNT signalling (Fig. 3g, h, Extended Data Fig. 8e–k, Supplementary Discussion). Together, these data suggest that class-2 mutations confer cistromic dominance and abolish TLE3-mediated repression of the WNT program of metastasis (Fig. 3i).

Class-3 rearrangements occur within the *PAX9* and *FOXA1* locus that is linearly conserved across the deuterostome superphylum[28] (Fig. 4a). Notably, almost all break ends were clustered within the *FOXA1* topologically associating domain (Extended Data Fig. 9a). We found that the genes located within the *FOXA1* topologically associating domain had the highest expression in the normal prostate, and the non-coding RP11-356O9.1 transcript had a prostate-specific expression (Extended Data Fig. 9b). Furthermore, in patient tumours, expression of RP11-356O9.1 was strongly correlated with *FOXA1* and *TTC6* expression (Extended Data Fig. 9c). Thus, to identify prostate-specific enhancers of the *FOXA1* topologically associating domain, we performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) and interrogated chromatin features in AR^{+} and AR^{-} prostate cells. Notably, a CTCF-bound intronic site in RP11-356O9.1 (hereafter denoted as *FOXMIND*) and a site within the 3′ untranslated region of *MIPOL1* were accessible and marked with active enhancer modifications only in AR^{+}FOXA1^{+} prostate cancer cells (Fig. 4b, Extended Data Fig. 9d). This strongly suggested that these conserved sites function as enhancer elements. Consistently, CRISPR knockout of these loci in VCaP cells led to a significant decrease in the expression of *FOXA1*

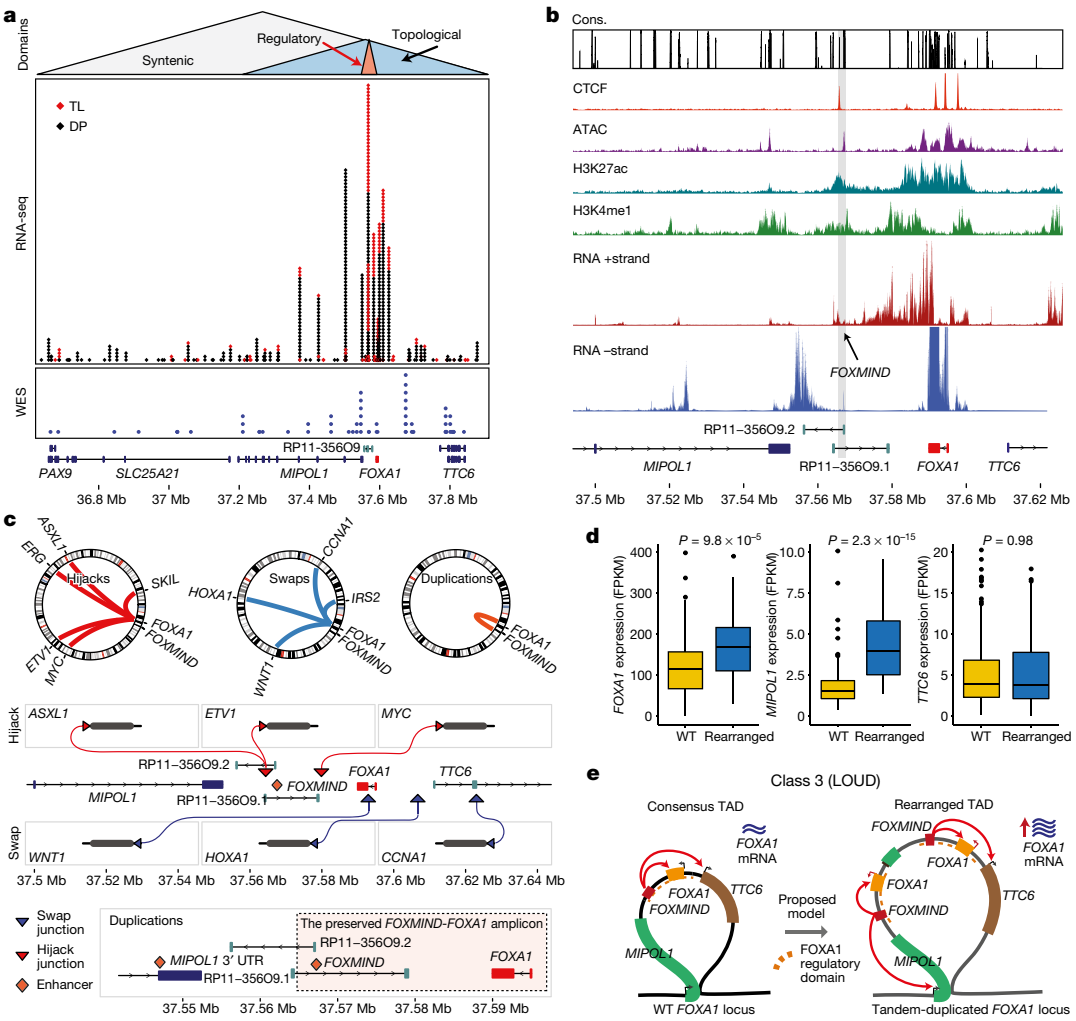


Fig. 4 | Genomic characterization of class-3 rearrangements of the *FOXA1* locus. a, Break ends in relation to the *FOXA1* syntenic, topological and regulatory domains. WES, whole-exome sequencing. **b**, Representative functional genomic tracks at the *FOXA1* locus. Base-level conservation (cons.), DNA accessibility (ATAC), enhancer-associated histone modifications (H3K27me1 and H3K27Ac), CTCF chromatin binding and stranded RNA-seq read densities are visualized. The *FOXMIND* enhancer is highlighted. **c**, Structural patterns of translocations and duplications. Hijacks occur between *FOXMIND* and *FOXA1*; swaps occur upstream of *FOXA1*. Duplications amplify the highlighted *FOXMIND-FOXA1* regulatory domain. **d**, Transcriptional changes in the *FOXA1*, *MIPOL1* and *TTC6* genes in wild-type ($n = 320$) and rearranged ($n = 50$) cases (two-sided *t*-test). Box plot centre, median; box, quartiles 1–3; whiskers, quartiles 1–3 ± 1.5 × IQR. FPKM, fragments per kilobase of transcript per million mapped reads. **e**, Class-3 model. Tandem duplications within the *FOXA1* topologically associating domain (TAD) amplify *FOXMIND* to drive overexpression of *FOXA1*.

and *TTC6*—but not of *MIPOL1*, which has its promoter outside of the *FOXA1* topologically associating domain (Extended Data Fig. 9d, e).

We found that translocations were largely within a 50-kb region between *FOXA1* and the 3′ untranslated region of *MIPOL1*, whereas break-end junctions from duplications mostly flanked the *FOXMIND-FOXA1* region (Fig. 4a, Extended Data Fig. 9f). For translocations, we delineated two patterns: (1) the hijacking of the *FOXMIND* enhancer; and (2) insertions upstream of the *FOXA1* promoter (Fig. 4c). The first pattern subsumes previously reported in-frame fusion genes that involve RP11-356O9.1, *ETV1*[29] and *SKIL*[30], as well as a newly reported *ASXL1* fusion (Supplementary Table 4). The second pattern inserts an oncogene (such as *CCNA1*) upstream of *FOXA1* (Fig. 4c). Notably, both mechanisms resulted in outlier expression of the translocated gene (Extended Data Fig. 9g). For duplications, which constitute 70% of all rearranged cases, we found *FOXMIND* and *FOXA1* to be co-amplified in 89% of the rearranged cases and never separated (Fig. 4c, bottom, Extended Data Fig. 9h), thus preserving the *FOXMIND-FOXA1* regulatory domain.

Next, while assessing the transcriptional effect of duplications, we found that levels of *FOXA1* mRNA were poorly correlated with copy number (Extended Data Fig. 10a), but highly sensitive to focal structural variants. Tandem duplications (ascertained at the RNA and DNA levels) significantly increased expression of *FOXA1* and *MIPOL1*, but not of *TTC6* (Fig. 4d). Translocations resulted in a modest decrease in expression levels of *FOXA1* (Extended Data Fig. 10b), despite a significant co-occurrence with tandem duplications (odds ratio = 3.89, Extended Data Fig. 10c). To investigate this further, we carried out haplotype-resolved, linked-read sequencing of MDA-PCA-2b cells, which contain a translocation of *FOXMIND* and *ETV1*. Here, *ETV1* translocation was accompanied by a focal tandem duplication in the non-translocated *FOXA1* allele (Extended Data Fig. 10d). The translocated *FOXA1* allele was inactivated, which resulted in monoallelic transcription (Extended Data Fig. 10e) without a net loss in *FOXA1* expression (266 fragments per kilobase of transcript per million mapped reads, 95th percentile in mCRPC). By contrast, RP11-356O9.1 retained biallelic expression (Extended Data Fig. 10f). In LNCaP cells, which also contain an *ETV1* translocation into the *FOXA1* locus, deletion of *FOXMIND* caused a significant reduction in *ETV1* expression (Extended Data Fig. 10g). Thus, translocations result in the loss of *FOXA1* expression from the allele *in cis*, which is rescued by tandem duplications of the allele *in trans*. Altogether, we propose a coalescent model in which class-3 structural variants duplicate or reposition

FOXMIND to drive overexpression of *FOXA1* or other oncogenes (Fig. 4e).

In summary, we identify three structural classes of FOXA1 alterations that differ in genetic associations and oncogenic mechanisms. We establish *FOXA1* as a principal oncogene in AR-dependent prostate cancer that is altered in 34.6% of mCRPC. Given the unique pathogenic features of the three classes, we have named them the 'FAST' (class-1), 'FURIOUS' (class-2) and 'LOUD' (class-3) alterations of FOXA1 (Figs. 2h, 3i, 4e, Supplementary Table 5, Supplementary Discussion). Structurally equivalent FOXA1 alterations are also found in other hormone-receptor-driven cancers, thus positioning FOXA1 as a promising target for therapeutic strategies in these malignancies.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1347-4.

Received: 22 May 2018; Accepted: 3 June 2019;
Published online 26 June 2019.

1. Gao, N. et al. Forkhead box A1 regulates prostate ductal morphogenesis and promotes epithelial cell maturation. *Development* **132**, 3431–3443 (2005).
2. Friedman, J. R. & Kaestner, K. H. The Foxa family of transcription factors in development and metabolism. *Cell. Mol. Life Sci.* **63**, 2317–2328 (2006).
3. Cancer Genome Atlas Research Network. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
4. Robinson, D. et al. Integrative clinical genomics of advanced prostate cancer. *Cell* **161**, 1215–1228 (2015).
5. Cancer Genome Atlas Research Network. The molecular taxonomy of primary prostate cancer. *Cell* **163**, 1011–1025 (2015).
6. Ciriello, G. et al. Comprehensive molecular portraits of invasive lobular breast cancer. *Cell* **163**, 506–519 (2015).
7. Dalin, M. G. et al. Comprehensive molecular characterization of salivary duct carcinoma reveals actionable targets and similarity to apocrine breast cancer. *Clin. Cancer Res.* **22**, 4623–4633 (2016).
8. Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
9. Jin, H.-J., Zhao, J. C., Ogden, I., Bergan, R. C. & Yu, J. Androgen receptor-independent function of FoxA1 in prostate cancer metastasis. *Cancer Res.* **73**, 3725–3736 (2013).
10. Jin, H.-J., Zhao, J. C., Wu, L., Kim, J. & Yu, J. Cooperativity and equilibrium with FOXA1 define the androgen receptor transcriptional program. *Nat. Commun.* **5**, 3972 (2014).
11. Song, B. et al. Targeting FOXA1-mediated repression of TGF-β signaling suppresses castration-resistant prostate cancer progression. *J. Clin. Invest.* **129**, 156–582 (2019).
12. Robinson, J. L. L. et al. Androgen receptor driven transcription in molecular apocrine breast cancer is mediated by FoxA1. *EMBO J.* **30**, 3019–3027 (2011).
13. Robinson, J. L. L. et al. Elevated levels of FOXA1 facilitate androgen receptor chromatin binding resulting in a CRPC-like phenotype. *Oncogene* **33**, 5666–5674 (2014).
14. Pomerantz, M. M. et al. The androgen receptor cistrome is extensively reprogrammed in human prostate tumorigenesis. *Nat. Genet.* **47**, 1346–1351 (2015).
15. Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
16. Iwafuchi-Doi, M. et al. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91 (2016).
17. Lupien, M. et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970 (2008).
18. Barbieri, C. E. et al. Exome sequencing identifies recurrent *SPOP*, *FOXA1* and *MED12* mutations in prostate cancer. *Nat. Genet.* **44**, 685–689 (2012).
19. Yang, Y. A. & Yu, J. Current perspectives on FOXA1 regulation of androgen receptor signaling and prostate cancer. *Genes Dis.* **2**, 144–151 (2015).
20. Grasso, C. S. et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239–243 (2012).
21. Gao, J. et al. 3D clusters of somatic mutations in cancer reveal numerous rare mutations as functional targets. *Genome Med.* **9**, 4 (2017).
22. Clark, K. L., Halay, E. D., Lai, E. & Burley, S. K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
23. Li, J. et al. Structure of the forkhead domain of FOXA2 bound to a complete DNA consensus site. *Biochemistry* **56**, 3745–3753 (2017).
24. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* **23**, 804–809 (2009).
25. Wang, Z. et al. BART: a transcription factor prediction tool with query gene sets or epigenomic profiles. *Bioinformatics* **34**, 2867–2869 (2018).
26. Behrens, J. et al. Functional interaction of β-catenin with the transcription factor LEF-1. *Nature* **382**, 638–642 (1996).
27. Daniels, D. L. & Weis, W. I. β-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat. Struct. Mol. Biol.* **12**, 364–371 (2005).
28. Wang, W., Zhong, J., Su, B., Zhou, Y. & Wang, Y.-Q. Comparison of *Pax1/9* locus reveals 500-Myr-old syntenic block and evolutionary conserved noncoding regions. *Mol. Biol. Evol.* **24**, 784–791 (2007).
29. Tomlins, S. A. et al. Distinct classes of chromosomal rearrangements create oncogenic ETS gene fusions in prostate cancer. *Nature* **448**, 595–599 (2007).
30. Annala, M. et al. Recurrent SKIL-activating rearrangements in ETS-negative prostate cancer. *Oncotarget* **6**, 6235–6250 (2015).

METHODS

Cell culture. Most cell lines were originally purchased from the American Type Culture Collection (ATCC) and were cultured as per standard ATCC protocols. LNCaP-AR and LAPC4 cells were gifts from the laboratory of C. Sawyers (Memorial Sloan Kettering Cancer Center). Unless otherwise stated, for all the experiments LNCaP, PNT2, LNCaP-AR, C42B, 22RV1, DU145 and PC3 cells were grown in the RPMI 1640 medium (Gibco) and VCaP cells in the DMEM with Glutamax (Gibco) medium supplemented with 10% full bovine serum (FBS; Invitrogen). LAPC4 cells were grown in IMEM (Gibco) supplemented with 15% FBS and 1 nM of R1881. For the immortalized normal prostate cells: RWPE1 cells were grown in keratinocyte medium with regular supplements (Lonza); PNT2 cells were grown in RPMI medium with 10% FBS. HEK293 cells were grown in DMEM (Gibco) medium with 10% FBS. All cells were grown in a humidified 5% CO_2 incubator at 37 °C. All cell lines were tested once a fortnight to be free of mycoplasma contamination and genotyped every month at the University of Michigan Sequencing Core using Profiler Plus (Applied Biosystems) and compared with corresponding short tandem repeat profiles in the ATCC database to authenticate their identity in culture between passages and experiments.

Antibodies. For immunoblotting, the following antibodies were used: FOXA1 N-terminal (Cell Signaling Technologies: 58613S; Sigma-Aldrich: SAB2100835); FOXA1 C-terminal (Thermo Fisher Scientific: PA5-27157; Abcam: ab23738); AR (Millipore: 06-680); LSD1 (Cell Signaling Technologies: 2139S); vinculin (Sigma Aldrich: V9131); H3 (Cell Signaling Technologies: 3638S); GAPDH (Cell Signaling Technologies: 3683); β-actin (Sigma Aldrich: A5316); β-catenin (Cell Signaling Technologies: 8480S); vimentin (Cell Signaling Technologies: 5741S); phospho(S33/S37/T41)-β-catenin (Cell Signaling Technologies: 8814S); LEF1 (Cell Signaling Technologies: 2230S); AXIN2 (Abcam: ab32197); and TLE3 (Proteintech: 11372-1-AP).

For co-immunoprecipitation and ChIP–seq experiments, the following antibodies were used: FOXA1 N-terminal (Cell Signaling Technologies: 58613S); FOXA1 C-terminal (Thermo Fisher Scientific: PA5-27157); AR (Millipore: 06-680); V5 tag (R960-25); and TLE3 (Proteintech: 11372-1-AP).

Immunoblotting and nuclear co-immunoprecipitation. Cell lysates were prepared using the RIPA lysis buffer (Thermo Fisher Scientific; cat. no. 89900) and denatured in the complete NuPage 1× LDS/reducing agent buffer (Invitrogen) with 10 min heating at 70 °C. Between 10 and 25 μg of total protein was loaded per well, separated on 4–12% SDS polyacrylamide gels (Novex) and transferred onto 0.45-μm nitrocellulose membrane (Thermo Fisher Scientific; cat. no. 88018) using a semi-dry transfer system (Trans-blot Turbo System; BioRad) at 25 V for 1 h. The membrane was incubated for 1 h in blocking buffer (Tris-buffered saline, 0.1% Tween (TBS-T), 5% non-fat dry milk) and incubated overnight at 4 °C with primary antibodies. When samples were run on multiple gels for an experiment, multiple loading control proteins (GAPDH, β-actin, total H3 and vinculin) were probed on each membrane separately. Host-species-matched secondary antibodies conjugated to horseradish peroxidase (HRP; BioRad) were used at 1/20,000 dilution to detect primary antibodies and blots were developed using enhanced chemiluminescence (ECL Prime, Thermo Fisher Scientific) following the manufacturer's protocol.

For nuclear co-immunoprecipitation assays, 8–10 million cells ectopically overexpressing different V5-tagged FOXA1 variants and wild-type AR (or TLE3) were fractionated to isolate intact nuclei using the NE-PER kit reagents (Thermo Fisher Scientific; cat. no. 78835) and lysed in the complete IP lysis buffer (Thermo Fisher Scientific; cat. no. 87788). Nuclear lysates were incubated for 2 h at 4 °C with 30 μl of magnetic protein-G Dynabeads (Thermo Fisher Scientific; cat. no. 10004D) for pre-clearing. A fraction of the pre-cleared lysate was saved as input and the remainder was incubated overnight (12–16 h) with 10 μg of target protein antibody at 4 °C with gentle mixing. Next day, 50 μl of Dynabeads protein-G beads were added to the lysate–antibody mixture and incubated for 2 h at 4 °C. Beads were washed three times with IP buffer (150 nM NaCl; Thermo Fisher Scientific) and directly boiled in 1× NuPage LDS/reducing agent buffer (ThermoFisher Scientific; cat. no. NP0007 and NP0009) to elute and denature the precipitated proteins. These samples were then immunoblotted as described above with the exception of using protein A-HRP secondary (GE HealthCare; cat. no. NA9120-1ML) antibody for detection.

RNA extraction and quantitative polymerase chain reaction. Total RNA was extracted using the the miRNeasy Mini Kit (Qaigen), with the inclusion of the on-column genomic DNA digestion step using the RNase-free DNase Kit (Qaigen), following the standard protocols. RNA was quantified using the NanoDrop 2000 Spectrophotometer (ThermoFisher Scientific) and 1 μg of total RNA was used for complementary DNA (cDNA) synthesis using the SuperScript III Reverse Transcriptase enzyme (Thermo Fisher Scientific) following the manufacturer's instructions. Twenty nanograms of cDNA was input per polymerase chain reaction (PCR) using the FAST SYBR Green Universal Master Mix (Thermo Fisher Scientific) and every sample was quantified in triplicate. Gene expression was calculated relative to *GAPDH* and *HPRT1* (loading control) using the $\Delta\Delta C_t$ method and normalized to the control group for graphing. Quantitative PCR (qPCR) primers were designed using the Primer3Plus tool (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) and synthesized by Integrated DNA Technologies.

Primer used in this study are listed below: *GAPDH*: forward (F), TGCACCACCA ACTGCTTAGC and reverse (R), GGCATGGACTGTGGTCATGAG; *HPRT1*: F, AGGCGAACCTCTCGGCTTTC and R, CTAATCACGACGCCAGGGCT; *ACTB*: F, AGGATGCAGAAGGAGATCACTG and R, AGTACTTGCGCTCAGGAGGAG; *AR*: F, CAGTGGATGGGCTGAAAAAT and R, GGAGCTTGGTGAGCTGGTAG; *FOXA1*-3′: F, GAAGACTCCAGCCTCCTCAACTG and R, TGCCTTGAAGTCCA GCTTATGC; *FOXA1*-5′: F, CTACTACGCAGACACGCAGG and R, CCGCTCGTAGTCATGGTGTT; *TLE3*: F, AAGGACAGCTTGAGCCGATA and R, TTTGGTCTTGGAGGAAGGTG; *TTC6*: F, CGAACAGAGCCAGGAGGT AG and R, GTTCTCCCTGGGCTCCTAAC; *MIPOL1*: F, GCAAACGGTTAGAGC AGGAG and R, GGGTCTGGATTTCCTCTTCC; *ETV1*: F, TACCCCATGGACC ACAGATT and R, CACTGGGTCGTGGTACTCCT; *TUBB*: F,CTGGACCGCATC TCTGTGTACT and R,GCCAAAAGGACCTGAGCGAACA.

siRNA-mediated gene knockdown. Cells were seeded in a 6-well plate at the density of 100,000–250,000 cells per well. After 12 h, cells were transfected with 25 nM of gene-targeting ON-TARGETplus SMARTpool siRNAs or non-targeting pool siRNAs as negative control (Dharmacon) using the RNAiMAX reagent (Life Technologies; cat. no. 13778075) on two consecutive days, following the manufacturer's instructions. Both total RNA and protein were extracted on day 3 (total 72 h) to confirm efficient (>80%) knockdown of the target genes. For crystal-violet staining, at day 9 growth medium was aspirated and cells were first fixed with 4% formaldehyde solution, followed by a 30-min incubation in 0.5% crystal-violet solution in 20% methanol, and then scanned. Catalogue numbers and guide sequences (5′ to 3′) of siRNA SMARTpools (Dharmacon) used are: non-targeting control (cat. no. D-001810-10-05; UGGUUUACAUGUCGACUAA, UGGUUUACAUGUUGUGUGA, UGGUUUACAUGUUUUCUGA, UGGUUUA CAUGUUUUCCUA); *AR* (cat. no. L-003400-00-0005; GAGCGUGGACUUUCCG GAA, UCAAGGAACUCGAUCGUAU, CGAGAGAGCUGCAUCAGUU, CAGAAAUGAUUGCACUAUU); *FOXA1* (cat. no. L-010319-00-0005; GCACUGCAAUACUCGCCUU, CCUCGGAGCAGCAGCAUAA, GAACAGCU ACUACGCAGAC, CCUAAACACUUCCUAGCUC); *TLE3* (cat. no. L-019929-00-0005; GCCAUUAUGUGAUGUACUA, GCAUGGACCCGAUAGGUAU, GAACCACCAUGAACUCGAU, UCAGGUCGAUGCCGGGUAA).

The *FOXA1* SMARTpool consists of siRNAs targeting 5′ as well as 3′ ends of the *FOXA1* transcript. Thus, both wild-type and class-2 mutant transcripts are degraded using the SMARTpool siRNAs. This was experimentally confirmed in LAPC4 cells that endogenously contain a *FOXA1* class-2 mutation (Extended Data Fig. 1d, e).

CRISPR–Cas9-mediated gene or enhancer knockout. Cells were seeded in a 6-well plate at the density of 200,000–300,000 cells per well and infected with viral particles with lentiCRISPR-V2 plasmids coding either non-targeting (sgNC) or single guide RNAs (sgRNAs) targeting the exon 1 or the FKHD of *FOXA1* (both resulting in FOXA1 inactivation). This was followed by three days of puromycin selection, after which proliferation assays were carried out as described below. The lentiCRISPR-V2 vector was a gift from the laboratory of F. Zhang (Addgene plasmid no. 52961).

sgRNA sequences used are as follows: sgNC no. 1: 5′-GTAGCGAACGTGTCC GGCGT-3′; sgNC no. 2: 5′-GACCGGAACGATCTCGCGTA-3′; sg*FOXA1* exon 1: 5′-GTAGTAGCTGTTCCAGTCGC-3′; sg*FOXA1* FKHD: 5′-GCCGTTCTCGAACATGTTGC-3′.

Alternatively, for functional interrogation of the *FOXA1* topologically associating domain (TAD) enhancer elements, VCaP or LNCaP cells were transfected with pairs of sgRNAs targeting the *MIPOL1* untranslated region (UTR) or *FOXMIND* or a control locus within the *FOXA1* TAD. Transfected cells were then selected with puromycin (1.0 μg/ml) for 48 h, followed by incubation for an additional 72 h. Total RNA was extracted and qPCR was performed as described above.

Pairwise sgRNA sequences are as follows (5′ to 3′): control sgRNA (sgCtrl): CA CCGATTAGCCTCAACTATACCA and CACCGTGCAATATCTGAATCACACG; sgMIPOL1 UTR: CACCGTGAAAAAAAACGACAGTCTG and CACCGAACTC AAGTCAGCAGCAAAG; sgFOXMIND 1: CACCGCTTTAATAAAGCTATTTGC and CACCGATAGAGTGACTAATGCCCTG; sgFOXMIND 2: CACCGTAACAGT TGACCTACTAAC and CACCGATTTAGATAAGGGGATAGAA; sgFOX-MIND 3: CACCGCTTTAATAAAGCTATTTGC and CACCGATTTAG ATAAGGGGATAGAA.

CRISPR knockout screen. For the genome-wide CRISPR knockout screen, a two-vector system was used. First, LNCaP cells were engineered to stably overexpress the enzymatically active Cas9 protein. These cells were then treated with the human GeCKO knockout sgRNA library (GecKO V2) that was a gift from the Zhang laboratory (Addgene; cat. no. 1000000049). This was followed by puromycin

selection for 48 h, after which a fraction of these cells was processed to isolated genomic DNA as the input sample. The remaining cells were then cultured for 30 days, and genomic DNA was extracted at this time point. sgRNA sequences were amplified using common adaptor primers and sequenced on the Illumina HiSeq 2500 (125-nucleotide read length). Sequencing data were analysed as described[31] and depletion or enrichment of individual sgRNAs at 30 days was calculated relative to the input sample. Note that only a subset of genes—including essential controls, epigenetic regulators and transcription factors from the GeCKO-V2 screen—was plotted in Extended Data Fig. 1i.

Proliferation assays. For siRNA growth assays, cells were directly plated in a 96-well plate at the density of 2,500–8,000 cells per well and transfected with gene-specific or non-targeting siRNAs, as described above, on day 0 and day 1. Every treatment was carried out in six independent replicate wells. CellTiter-Glo reagent (Promega) was used to assess cell viability at multiple time points after tranfection, following the manufacturer's protocol. Data were normalized to readings from siNC treatment on day 1, and plotted as relative cell viability to generate growth curves.

Alternatively, for CRISPR sgRNA growth assays, cells were treated as described above for target-gene inactivation and seeded into a 24-well plate at 20,000 cells per well, with 2 replicates per group. After 12 h, plates were placed into the IncuCyte live-cell imaging machine (IncuCyte) set at the phase-contrast option to record cell confluence every 3 h for between 7 and 9 days. Similarly, for class-1 growth assays (Fig. 2f), stable doxycycline-inducible 22RV1 cells were grown in 10% charcoal-stripped-serum (CSS)-supplemented medium for 48 h. Androgen-starved cells were then seeded into a 96-well plate at 5,000 cells per well in 10% CSS medium with or without addition of doxycycline (1 μg/ml) to induce control or mutant protein expression (6 replicates per group). Once adherent, treated cells were placed in the IncuCyte live-cell imaging machine set at phase contrast to record cell confluence every 3 h for between 7 and 9 days. In all IncuCyte assays, confluence measurements from all time points were normalized to the matched measurement at 0 h and plotted as relative confluence to generate growth curves.

Cloning of representative FOXA1 mutants. Wild-type *FOXA1* coding sequence was purchased from Origene (cat. no. SC108256) and cloned into the pLenti6/V5 lentiviral vector (Thermo Fisher Scientific; cat. no. K4955-10) using the standard TOPO cloning protocol. Class-1 missense mutations (I176M, H247Q and R261G) were engineered from the wild-type FOXA1 vector using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Tech) as per the manufacturer's instructions. All point mutations were confirmed using Sanger sequencing through the University of Michigan Sequencing Core Facility. Engineered mutant plasmids were further transfected in HEK293 cells to confirm expression of the mutant protein. For truncated class-2 variants, the wild-type coding sequence up to the amino acid before the intended mutation was cloned. All FOXA1 variants had the V5 tag fused on the C terminus. Selected mutants were cloned into a doxycycline-inducible vector (Addgene: pCW57.1; cat. no. 41393) to generate stable lines. For FRAP and single particle tracking assays, the pCW57.1 vector was edited to incorporate an in-frame GFP or Halo coding sequences at the C-terminal end, respectively.

FRAP assay and data quantification. PNT2 cells were seeded in a 6-well plate at 200,000 cells per well, and transfected with 2 μg of doxycycline-inducible vectors that encoded different variants of FOXA1 fused to GFP on the C-terminal end. After 24 h, cells were plated in glass-bottom microwell dishes (MatTek; #P35G-1.5-14-C) in phenol-free growth medium supplemented with doxycycline (1 μg/ml). Cells were then incubated for 48 h to allow for robust expression of the exogenous GFP-tagged protein and strong adherence to the glass surface. Microwell dishes were placed in humidity-controlled chamber set at 37 °C (Tokai-Hit) and mounted on the SP5 Inverted 2-Photon FLIM Confocal microscope (Leica). FRAP Wizard from the Leica Microsystems software suite was used to conduct and analyse FRAP experiments. Fluorescence signals were automatically computed in regions of interest using in-built tools in the FRAP Wizard. Roughly half of the nucleus was photobleached using the argon laser at 488 nm and 100% intensity for 20–30 iterative frames at 1.2-s intervals. Laser intensity was reduced to 1% for imaging post bleaching. Immediately after photobleaching, 2 consecutive images were collected at 1.2-s intervals followed by images taken at 10-s intervals for 60 frames (that is, 10 min).

For data analyses, recovery of signal in the bleached half and loss of signal in the unbleached half were measured as average fluorescence intensities in at least 80% of the respective areas, excluding the immediate regions flanking the separating border. All intensity curves were generated from background-subtracted images. The fluorescence signal measured in a region of interest was normalized to the signal before bleaching using the following formula[32]: $R = (I_t - I_{bg})/(I_o - I_{bg})$, in which I_o is the average intensity in the region of interest before bleaching, I_t is the average intensity in the region of interest at any time-point after bleaching and I_{bg} is the background fluorescence signal in a region outside of the cell nucleus. Raw recovery kinetic data from above were fitted with best hyperbolic curves using the GraphPad Prism software and the time until 50% recovery was calculated from the resulting best-fit equations. For representative time-lapse nuclei images shown in the FRAP figures, the fluorescence signal was uniformly brightened for ease of visualization.

Single particle tracking experiment and data quantification. PNT2 cells were transiently transfected with doxycycline-inducible vectors encoding C-terminal Halo-tagged wild-type or class-1 mutant variants of FOXA1. Transfected cells were seeded in glass-bottomed DeltaT culture dishes (Bioptechs; cat. no. 04200417C) and incubated for 24 h with 0.01 μg/ml of doxycycline. Cells were then treated with phenol-red-free medium containing 2% FBS and 5 nM cell permeable JF549 Halo ligand dye[33] for 30 min at 37 °C. Cells were subsequently washed twice, 10 min per wash at 37 °C, with phenol-red-free medium containing 2% FBS. Before imaging, cells were washed once with the 1× HBSS buffer and were imaged in the buffer.

Single particle tracking was performed on an Olympus IX81 microscope via HILO illumination, as previously described[34], at a spatial accuracy of 30 nm and temporal resolution of 33 ms. Image analysis was performed as previously described[35]. In brief, tracking was done in Imaris (bitplane) and particles that were at least visible for four continuous frames were used for further analysis. Diffusion constants were calculated as previously described[36], assuming a Brownian diffusion model under steady-state conditions. Dwell time histograms were fit to a double-exponential function to extract fast and slow dwell times of 'bound' particles that displayed a frame-to-frame displacement of <300 nm. All particles that were visible for less than 4 consecutive frames, or those that moved >300 nm between frames, were counted as 'unbound' particles. At least 5 cells were imaged for each transcription factor variant and >500 particles were tracked to extract diffusion constants and dwell time.

Dual luciferase AR reporter assay. HEK293 cells stably overexpressing the wild-type AR protein (that is, HEK293-AR) were used for the AR reporter assays. HEK293-AR cells were seeded in a 12-well plate at 300,000 cells per well and transfected with 2 μg of the pLenti6/V5 vector encoding different variants of FOXA1, or GFP (control). After 8 h, medium was replaced with 10% CSS-supplemented phenol-free medium (androgen-depleted) and cells were transfected with the AR reporter Firefly luciferase or negative-control constructs from the Cignal AR-Reporter(luc) kit (Qiagen; cat. no. CCS-1019L) as per the manufacturer's instructions. Both constructs were premixed with constitutive *Renilla* luciferase vector as control. After 12 h, cells were treated with different dosages of DHT or enzalutamide (at 10 μM dosage); and additional 24 h later dual luciferase activity was recorded for every sample using the Dual-Glo Luciferase assay (Promega; E2980) and luminescence plate reader (Promega-GLOMAX-Multi Detection System). Each treatment condition had four independent replicates. Firefly luciferase signals were normalized with the matched *Renilla* luciferase signals to control for variable cell number and/or transfection efficiencies, and normalized signals were plotted relative to the negative control reporter constructs.

Electrophoretic mobility shift assay. HEK293 cells were plated in 10-cm dishes at 1 million per plate and transfected with 10 μg of the pLenti6/V5 vector coding GFP (control) or different variants of FOXA1. After 48 h, cells were trypsinized and nuclear lysates were prepared using the NE-PER kit reagents (Thermo Fisher Scientific). Immunoblots were run to confirm comparable expression of recombinant FOXA1 variants in 2 μl (that is, equal volume) of final nuclear lysates. Next, FOXA1 and AR ChIP–seq data were used to identify the *KLK3* enhancer element. Sixty base pairs of the *KLK3* enhancer, centred at the *FOXA1* consensus motif 5′-GTAAACAA-3′, were synthesized as single-stranded oligonucleotides (IDT) and biotin-labelled using the Biotin 3′-End DNA labelling kit (Thermo Fisher Scientific), and then annealed to generate a labelled double-stranded DNA duplex.

Binding reactions were carried out in 20-μl volumes containing 2 μl of the nuclear lysates, 50 ng/μl poly(dI.dC), 1.25% glycerol, 0.025% Nonidet P-40 and 5 mM $MgCl_2$. Biotin-labelled *KLK3* enhancer probe (10 fmol) was added at the very end with gentle mixing. Reactions were incubated for 1 h at room temperature, size-separated on a 6% DNA retardation gel (100 V for 1 h; Invitrogen) in 0.5× TBE buffer, and transferred on the Biodyne Nylon membrane (0.45 μm; Thermo Fisher Scientific) using a semi-dry system (BioRad). Transferred DNA was crosslinked to the membrane using the UV light at 120 mJ/cm^2 for 1 min. Biotin-labelled free and protein-bound DNA was detected using HRP-conjugated streptavidin (Thermo Fisher Scientific) and developed using chemiluminescence according to the manufacturer's protocol.

Protein synthesis and purification. First, wild-type FOXA1 and FOXA1(P358fs) proteins were purified using the *Escherichia coli* bacterial expression system and nickel-affinity chromatography. In brief, wild-type FOXA1 or FOXA1(P358fs) coding sequences were cloned into the pFC7A (HQ) Flexi vector (Promega; cat. no. C8531) with a C-terminal HQ tag, following the manufacturer's protocol. These expression constructs were used to transform Single Step (KRX) Competent *E. coli* cells (Promega; cat. no. L3002), which have been modified for synthesis of mammalian proteins. A starter broth of 2 ml was inoculated with a single colony of transformed bacterial cells and incubated at 37 °C with constant shaking at

250 rpm until an optical density at 600 nm (OD_{600}) of 0.4–0.5 was reached. The starter brother was then used to inoculate 1,000 ml of LB broth containing ampicillin, and protein synthesis was induced using 0.1% v/v of rhamanose. Induced culture was incubated at 20 °C for 16 h with constant shaking at 250 rpm. Bacterial cells were then pelleted by centrifugation at 4,000 rpm for 30 min and mechanically lysed through sonication in 50 mM Tris (pH 7.4), 150 mM NaCl, 1 mM $MgCl_2$, 0.5 mM EDTA, 1 mM DTT and 1% glycerol in the presence of protease inhibitors (Roche). HisLink Purification Resin (Promega; cat. no. V8821) was used to purify untagged recombinant proteins from the crude bacterial lysates as per the manufacturer's protocol (this also includes removal of the His tag). Purified protein fractions were then tested for purity by Coomaisse staining relative to the crude input lysates, and purified protein concentrations were estimated using protein standards of known concentrations (Thermo Fisher Scientific; cat. no. 23208). The identities of purified proteins were confirmed via immunoblotting using an N-terminal FOXA1 antibody (Cell Signaling Technology; cat. no. 58613S).

Biolayer interferometry assay. Biolayer interferometry (BLI) assays were carried out using the Octet-RED96 system (PALL ForteBio) and in-built analysis software. In brief, a biotin-labelled, 60-bp *KLK3* enhancer element centred at the FOXA1 consensus motif was immobilized on the Super Streptavidin Biosensors (PALL ForteBio, part no. 18-5057) with the loading step carried out for 1,000 s with shaking at 500 rpm. This was followed by baseline measurements for 120 s and association for 900 s using varying concentrations of purified FOXA1 proteins (3.125–100 nM; two replicate biosensors per concentration). A control DNA element with no FOXA1 motif was used in the negative-control reaction with 100 nM of the protein. The association step was followed by the dissociation step for 3,000 s. Signal from all the biosensors was adjusted for the background signal from the control sensors and normalized data of DNA binding kinetics were analysed using the Octet-RED96 (PALL ForteBio) analysis software, as previously described[37].

Generation of CRISPR clones and stable lines. 22RV1 or LNCaP cells were seeded in a 6-well plate at 200,000 cells per well and transiently transfected with 2.5 μg of lentiCRISPR-V2 (Addgene; 52961) vector using the Lipofectamine 3000 reagent (cat. no. L3000008), encoding the Cas9 protein and sgRNA that cuts either at amino acid 271 (5′-GTCAAGTGCGAGAAGCAGCCG-3′) or 359 (5′-GCCGGGCCCGGAGCTTATGGG-3′) in exon 2 of *FOXA1*. Cells were treated with non-targeting control sgRNA (5′-GACCGGAACGATCTCGCGTA-3′) vector to generate isogenic wild-type clones. Transfected cells were selected with puromycin (Gibco) for 3–4 days and sorted by fluorescence-activated cell sorting as single cells into 96-well plates. Cells were maintained in 96-well plates for 4–6 weeks, with replacement of the growth medium every 7 days to allow for the expansion of clonal lines. Clones that successfully seeded were further expanded and genotyped for *FOXA1* using Sanger sequencing, and immunoblotting with the N-terminal FOXA1 antibody. Sequence- and expression-validated 22RV1 and LNCaP clones with distinct class-2 mutations were used for growth, invasion and metastasis assays as described.

To generate stable cells, doxycycline-inducible vectors coding different variants of FOXA1 or GFP (control) were packaged into viral particles at the University of Michigan Vector Core. Prostate cancer cells were seeded in a 6-well plate at 100,000–250,000 cells per well and infected with 0.5 ml of 10× viral titres packaged at the University of Michigan Vector Core. This was followed by 3–4 days of puromycin (Gibco) selection to generate stable lines.

Rescue growth and functional compensation experiments. Stable 22RV1 cells with doxycycline-inducible expression of empty vector (control), FOXA1 wild type, or distinct FOXA1 mutants were seeded in a 6-well plate in the completed growth medium supplemented with 1.0 μg/ml of doxycycline. Notably, the exogenous genes only contain the coding sequence of *FOXA1* without its intron and UTRs. After 24 h, cells were transfected with 30 nM of either distinct 3′ UTR-specific *FOXA1*-targeting siRNAs or a non-targeting control siRNA using the RNAiMAX (Life Technologies; cat. no. 13778075) reagent. *FOXA1* UTR-specific siRNAs were purchased from Thermo Fisher Scientific (cat. no. siNC, 4390844 (sequence is proprietary); siRNA no. 3, s6687 (sense sequence: 5′-GCAAUACUCUUAACCAUAA-3′); siRNA no. 4, 5278 (sense sequence: 5′-AACACATAAAATTAGTTTC-3′); and siRNA no. 5 – 107428 (sense sequence: 5′-AAGTTATAGGGAGCTGGAT-3′)). On the following day, cells were counted and seeded in a 96-well plate at a density of 5,000 cells per well with 6 replicates for each treatment condition. Cell growth was then assessed using the IncuCyte assay, as described above.

Testing the GFP-tagged wild-type FOXA1 variant. 22RV1 cells were seeded in 10-cm dishes and transfected with 8 μg of mammalian expression plasmids encoding either FOXA1(WT) or FOXA1(WT)–GFP (the exact construct used in the FRAP assay) using the Lipofectamine 3000 (Life Technologies; cat. no. L3000008) reagent, as per the manufacturer's protocol. Transgene expression was induced using 1.0 μg/ml of doxycycline and cells were cultured for 96 h with doxycycline replenishment every 48 h. Total RNA was extracted and RNA-seq was performed as described. A portion of these cells was used for the rescue growth experiments using UTR-specific *FOXA1* siRNAs as described above.

Matrigel invasion assay. 22RV1 CRISPR clones were grown in 10% CSS-supplemented medium for 48 h for androgen starvation. A matrigel-coated invasion chamber was used, which was additionally coated with a light-tight polyethylene terephthalate membrane to allow for fluorescent quantification of the invaded cells (Biocoat: 24-well format, no. 354166). Fifty thousand starved cells were resuspended in serum-free medium and were added to each invasion chamber. Twenty per cent FBS-supplemented medium was added to the bottom wells to serve as a chemoattractant. After 12 h, medium from the bottom well was aspirated and replaced with 2 μg/ml Calcein-green AM dye (Thermo Fisher Scientific; C3100MP) in 1× HBSS (Gibco) and incubated for 30 min at 37 °C. Invasion chambers were then placed in a fluorescent plate reader (Tecan-Infinite M1000 PRO) and fluorescent signals from the invaded cells at the bottom were averaged across 16 distinct regions per chamber to determine the extent of invasion.

ChIP–seq. ChIP experiments were carried out using the HighCell# ChIP-Protein G kit (Diagenode) as per the manufacturer's protocol. Chromatin from five million cells was used per ChIP reaction with 6.5 μg of the target protein antibody. In brief, cells were trypsinized and washed twice with 1× PBS, followed by crosslinking for 8 min in 1% formaldehyde solution. Crosslinking was terminated by the addition of 1/10 volume 1.25 M glycine for 5 min at room temperature followed by cell lysis and sonication (Bioruptor, Diagenode), resulting in an average chromatin fragment size of 200 bp. Fragmented chromatin was then used for immunoprecipitation using various antibodies, with overnight incubation at 4 °C. ChIP DNA was de-crosslinked and purified using the iPure Kit V2 (Diagenode) using the standard protocol. Purified DNA was then prepared for sequencing as per the manufacturer's instructions (Illumina). ChIP samples (1–10 ng) were converted to blunt-ended fragments using T4 DNA polymerase, *E. coli* DNA polymerase I large fragment (Klenow polymerase) and T4 polynucleotide kinase (New England BioLabs (NEB)). A single A base was added to fragment ends by Klenow fragment (3′ to 5′ exo minus; NEB) followed by ligation of Illumina adaptors (Quick ligase, NEB). The adaptor-ligated DNA fragments were enriched by PCR using the Illumina Barcode primers and Phusion DNA polymerase (NEB). PCR products were size-selected using 3% NuSieve agarose gels (Lonza) followed by gel extraction using QIAEX II reagents (Qiagen). Libraries were quantified and quality checked using the Bioanalyzer 2100 (Agilent) and sequenced on the Illumina HiSeq 2500 Sequencer (125-nucleotide read length).

Zebrafish embryo metastasis experiment. Wild-type AB^{TL} zebrafish were maintained in aquaria according to standard protocols. Embryos were generated by natural pairwise mating and raised at 28.5 °C on a 14 h light/10 h dark cycle in a 100-mm Petri dish containing aquarium water with methylene blue to prevent fungal growth. All experiments were performed with 2–7-day-old embryos post-fertilization, and were done in approved University of Michigan fish facilities using protocols approved from the University of Michigan Institutional Animal Care and Use Committee (UM-IACUC). Cell injections were carried out as previously described[38]. In brief, GFP-expressing normal (control) or cancer cells were resuspended in PBS at the concentration of 1×10^7 cells/ml. Forty-eight hours after fertilization, wild-type embryos were dechorionated and anaesthetized with 0.04 mg/ml tricaine. Approximately 10 nl (approximately 100 cancer cells) were microinjected into the perivitelline space using a borosilliac micropipette tip with filament. Embryos were returned to aquarium water and washed twice to remove tricaine, then moved to a 96-well plate with one embryo per well and kept at 35 °C for the duration of the experiment. All embryos were imaged at 24-h intervals to follow metastatic dissemination of injection cells. Water was changed daily to fresh aquarium water. More than 30 fish were injected for each condition (wild-type no. 2, $n = 30$; wild-type no. 5, $n = 50$; no. 57, $n = 35$; no. 84, $n = 57$; no. 113, $n = 38$) and metastasis was visually assessed daily up to 5 days after injection (that is, for a total of 7 days post-fertilization) by counting the total number of distinct cellular foci in the body of the embryos. All of the metastasis studies were terminated at seven days post-fertilization in accordance with the approved embryo protocols. Embryos were either imaged directly in the 96-well plates or placed onto a concave glass slide to capture representative images using a fluorescent microscope (Olympus-IX71). For quantification, evidently distinct cell foci in the embryo body were counted 72 h after the injections.

For all these experiments, relevant ethical regulations were carefully followed. No statistical methods were used to predetermine sample size for any of the cohort analyses or experiments. The experiments were not randomized and investigators were not blinded to allocation during experiments and outcome assessment unless otherwise stated.

ATAC-seq and data analysis. ATAC-seq was performed as previously described[39]. In brief, 25,000 normal prostate or prostate cancer cells were washed in cold PBS and resuspended in cytoplasmic lysis buffer (CER-I from the NE-PER kit, Invitrogen, cat. no. 78833). This single-cell suspension was incubated on ice for 10 min with gentle mixing by pipetting at every 2 min. The lysate was centrifuged at 1,300*g* for 5 min at 4 °C. Nuclei were resuspended in 2× TD buffer, then incubated with Tn5 enzyme for 30 min at 37 °C (Nextera DNA Library Preparation Kit;

cat. no. FC-121-1031). Samples were immediately purified by Qiagen minElute column and PCR-amplified with the NEBNext High-Fidelity 2X PCR Master Mix (NEB; cat. no. M0541L). qPCR was used to determine the optimal PCR cycles to prevent over-amplification. The amplified library was further purified by Qiagen minElute column and SPRI beads (Beckman Coulter; cat. no. A63881). ATAC-seq libraries were sequenced on the Illumina HiSeq 2500 (125-nucleotide read length).

Paired-end .fastq files were uniquely aligned to the hg38 human genome assembly using Novoalign (Novocraft) (with the parameters -r None -k -q 13 -k -t 60 -o sam –a CTGTCTCTTATACACATCT), and converted to .bam files using SAMtools (version 1.3.1). Reads mapped to mitochondrial or duplicated reads were removed by SAMtools and PICARD MarkDuplicates (version 2.9.0), respectively. Filtered .bam files from replicates were merged for downstream analysis. MACS2 (2.1.1.20160309) was used to call ATAC-seq peaks. The coverage tracks were generated using the program bam2wig (http://search.cpan.org/dist/Bio-ToolBox/) with the following parameters:–pe–rpm–span–bw. Bigwig files were then visualized using the IGV (Broad Institute) open source genome browser.

ChIP–seq data analysis. Paired-end 125-bp reads were trimmed and aligned to the GRCh38 human reference using the STAR (version 2.4.0g1) aligner with splicing disabled; the resulting reads were filtered using samtools 'samtools view -@ 8 -S -1 -F 384'. The resulting .bam file was sorted and duplicate-marked using Novosort, and converted into a bigwig file for visualization using 'bedtools genomecov -bg -split -ibam' and 'bedGraphToBigWig'. The coverage signal was normalized to total sequencing depth/1 $\times$ 10^6 reads. Peak calling was performed using MACS2 with the following settings: 'macs2 callpeak–call-summits–verbose 3 -g hs -f BAM -n OUT–qvalue 0.05'. ChIP peak profile plots and read-density heat maps were generated using deepTools2[40], and cistrome overlap analyses were carried out using the ChIPpeakAnno[41] package in R. It is important to note that, given the cistromic dominance of class-2 mutants, in heterozygous class-2 mutant clones part of the FOXA1 antibody binds to the wild-type protein that does not interact with, or immunoprecipitate, the DNA. This confounds all analyses involving peak-read density comparisons between the wild-type and class-2-mutant FOXA1 ChIP–seq data; we therefore largely avoided this strategy in our study. For the same reason, the read densities from only the heterozygous clones were factored by 1.5 for heat map generation in Fig. 3d.

De novo and known motif enrichment analysis. All de novo and known motif enrichment analyses were performed using the HOMER (v.4.10) suite of algorithms[42]. Peaks were called by the findPeaks function (-style factor -o auto) at 0.1% false discovery rate; de novo motif discovery and enrichment analysis of known motifs were performed with findMotifsGenome.pl (-size 200 -mask). For motif analysis of common wild-type- and mutant-specific chromatin binding sites, the top 5,000 peaks ranked by score were used as input. A common set of background sequences was generated by di-nucleotide shuffling of the input sequences using the fasta-shuffle-letters function from MEME[43]. Alternatively, we ranked peaks by the relative signal fold change between mutant and wild type, and selected the top and bottom 5,000 peaks (keeping the requirement that mutant-specific peaks are not called in the wild-type cistrome, and vice versa) for motif discovery. For class-2 mutants, only heterozygous 22RV1 clones were used, which more accurately recapitulate the clinical presentation of *FOXA1* mutations. Also, for both mutational classes, cistromes from biological replicates were merged to define a union cistrome that was compared to the union wild-type cistrome generated from matched FOXA1 wild-type cells. For the supervised motif analyses, we identified all instances of the FOXA canonical motif (5′-T[G/A]TT[T/G]AC-3′) within cistromes (ChIP–seq peaks) of class-1 and wild-type FOXA1 proteins using motifmatchR, and calculated nucleotide frequencies in the flanking positions.

Cohorts, datasets and resources. This study uses previously published public or restricted patient genetic data. Genetic calls for primary prostate cancer and breast cancer were obtained from the Genomic Data Commons (GDC)[44] for the prostate cancer PRAD[5] and breast cancer BRCA[6,45] cohorts, respectively. Raw RNA-seq data (paired-end reads from unstranded polyA libraries) for the samples were downloaded from the GDC and processed with our standard clinical RNA-seq pipeline CRISPR/CODAC (see below). For The Cancer Genome Atlas (TCGA) PRAD and BRCA cohorts, we downloaded mutational calls from multiple sources (GDC, cBio Portal and UCSC Xena) and additionally used the BAM-slicing tool to download sequence alignments from whole-exome sequencing libraries to the *FOXA1* locus. We then used our internal pipeline (see below) to call single-nucleotide variants and indels within *FOXA1*. We also used the downloaded aligned data for manual review of *FOXA1* mutation calls. Mutation calls for advanced primary and metastatic cases were obtained from the MSK-IMPACT cohort (downloaded from the cBio portal[46]). The main MCTP mCRPC cohort includes 360 previously reported cases (the location of all raw .bam files is provided in ref. [47]), the 10 additional mCRPC cases included here (but not in ref. [47]) will be included in the Database of Genotypes and Phenotypes (dbGaP) under accession code phs000673.v3.p1, and belong to a continuous sequencing program with the same IRB-approved protocol (MI-Oncoseq program, University of Michigan Clinical Sequencing Exploratory Research). The genetic sequencing data (WXS) for rapid autopsy cases are available from dbGaP with accession codes hs000554.v1.p1and phs000567.v1.p1. De-identified somatic mutation calls, RNA-seq fusion calls, processed and segmented copy-number data, and RNA-seq expression matrices across the full 370 cases of the MCTP mCRPC cohort are available on request from the authors.

Preparation of whole-exome sequencing and RNA-seq libraries. Integrative clinical sequencing (comprising exome sequencing and polyA and/or capture RNA-seq) was performed using standard protocols in our Clinical Laboratory Improvement Amendments-compliant sequencing laboratory. In brief, tumour genomic DNA and total RNA were purified from the same sample using the AllPrep DNA/RNA/miRNA kit (Qiagen). Matched normal genomic DNA from blood, buccal swab or saliva was isolated using the DNeasy Blood & Tissue Kit (Qiagen). RNA-seq was performed using the exome-capture transcriptome platform[48]. Exome libraries of matched pairs of tumour and normal DNA were prepared as previously described[49], using the Agilent SureSelect Human All Exon v4 platform (Agilent). All the samples were sequenced on an Illumina HiSeq 2000 or HiSeq 2500 (Illumina) in paired-end mode. The primary base call files were converted into FASTQ sequence files using the bcl2fastq converter tool bcl2fastq-1.8.4 in the CASAVA 1.8 pipeline.

Analysis of whole-exome sequencing data. The .fastq sequence files from whole-exome libraries were processed through an in-house pipeline constructed for analysis of paired tumour and normal data. The sequencing reads were aligned to the GRCh37 reference genome using Novoalign (version 3.02.08) (Novocraft) and converted into .bam files using SAMtools (version 0.1.19). Sorting, indexing, and duplicate marking of .bam files used Novosort (version 1.03.02). Mutation analysis was performed using freebayes (version 1.0.1) and pindel (version 0.2.5b9). Variants were annotated to RefSeq (via the UCSC genome browser, retrieved on 22 August 2016), as well as COSMIC v.79, dbSNP v.146, ExAC v.0.3 and 1000 Genomes phase 3 databases using snpEff and snpSift (v.4.1g). Single nucleotide variants and indels were called as somatic if they were present with at least 6 variant reads and 5% allelic fraction in the tumour sample, and present at no more than 2% allelic fraction in the normal sample with at least 20$\times$ coverage. Additionally, the ratio of variant allelic fractions between tumour and normal samples was required to be at least six to avoid sequencing and alignment artefacts at low allelic fractions. Minimum thresholds were increased for indels observed to be recurrent across a pool of hundreds of platform- and protocol-matched normal samples. Specifically, for each such indel, a logistic regression model was used to model variant and total read counts across the normal pool using PCR duplication rate as a covariate, and the results of this model were used to estimate a predicted number of variant reads (and therefore allelic fraction) for this indel in the sample of interest, treating the total observed coverage at this genomic position as fixed. The variant read count and allelic fraction thresholds were increased by these respective predicted values. This filter eliminates most recurrent indel artefacts without affecting our ability to detect variants in homopolymer regions from tumours exhibiting microsatellite instability. Germline variants were called using 10 variant reads and 20% allelic fraction as minimum thresholds, and were classified as rare if they had less than 1% observed population frequency in both the 1000 Genomes and ExAC databases. Exome data were analysed for copy-number aberrations and loss of heterozygosity by jointly segmenting B-allele frequencies and $\log_2$-transformed tumour/normal coverage ratios across targeted regions using the DNAcopy (version 1.48.0) implementation of the Circular Binary Segmentation algorithm. The expectation–maximization algorithm was used to jointly estimate tumour purity and classify regions by copy-number status. Additive adjustments were made to the $\log_2$-transformed coverage ratios to allow for the possibility of non-diploid tumour genomes; the adjustment resulting in the best fit to the data using minimum mean-squared error was chosen automatically and manually overridden if necessary.

Detection of copy-number break ends from whole-exome sequencing. The output of our clinical whole-exome sequencing pipeline includes segmented copy-number data, inferred absolute copy numbers and predicted parent-specific genotypes (for example, AAB), detection of loss of heterozygosity, and detection of copy-neutral loss of heterozygosity (uniparental disomy). Together, these data enable the detection of joint discontinuities in the copy-number profile (log-ratio and B-allele frequencies) at exon-level resolution. A subset of genomic rearrangements results in changes in copy number or allelic shifts, and the presence of such discontinuities in paired tumour-normal whole-exome sequencing data are therefore strongly indicative of a somatic breakpoint. For example, one copy gain will result in a segment with an increased log-ratio, and a corresponding zygosity deviation (see above). This segment will be discontinuous with adjacent segments, which will result in the call of a whole-exome sequencing break end (discontinuity) on either side of the copy gain. The size of the break end depends on the density of covered exons and in general the resolution is better in genic versus intergenic regions. We assessed the presence of such breakpoints within the gene-dense and exon-dense *FOXA1* locus; all copy-number break ends met statistical thresholds of the circular binary segmentation (CBS) algorithm (see above) at either the log-ratio or B-allele level.

Genetic characterization of mCRPC tumour samples at the pathway level. The co-occurrence or mutual exclusivity of *FOXA1* alterations with other previously described genetic events in prostate cancer has been carried out at the pathway level, but grouping putative functionally equivalent (and largely genetically mutually exclusive) events. All known types of ETS fusion (*ERG*, *ETV1*, *FLI1*, *ETV4* and *ETV5*) were considered as ETS-positive tumours, PI3K alterations included *PTEN* homozygous loss, *PIK3CA* activating mutations and *PIK3R1* inactivating mutations, AR pathway alterations included *AR*, *NCOR1*, *NCOR2* and *ZBTB16* mutations or deletions, but excluded *AR* amplifications and copy gains. The KMT category included mutations in all recurrently mutated lysine methyltransferases. The WNT category included inactivating alterations in *APC* and activating mutations in *CTNNB1*. DRD included cases with mutations in *BRCA1*, *BRCA2*, *PALB2* and *ATM* (all common mismatch repair genes), and *CDK12*.

Assessment of two-hit biallelic alterations. To assess the frequency of genetic inactivations of both alleles we integrated mutational, copy-number and RNA-seq (fusion) data. A gene was considered as having both alleles inactivated for any combination (pair) of the following events: copy loss, mutation, truncating fusion and copy-number breakpoint, in addition to homozygous deletion of both copies and two independent mutations. Ambiguous cases were manually reviewed to increase the accuracy and ascertain whether both events, for example, copy-number breakpoint and gene fusion, are probably independent events.

Unified mutation calling and variant classification of FOXA1. Mutation calls for *FOXA1* obtained or downloaded from the GDC and TCGA flagship manuscripts[5,6] as well as our internal pipelines were lifted over to GRCh38 (using the Bioconductor package rtracklayer) and annotated with respect to the canonical RefSeq *FOXA1* isoform. For TCGA samples or cases, multiple call sets were available and we manually reviewed all discrepancies in *FOXA1* mutation calls, resulting in a unified call set with improved sensitivity and specificity. Mutational effect (consequence) was simplified into three categories: missense, in-frame indel and frameshift (the last category included stop-gain, stop-loss and splice-site mutations). The resulting mutations were dichotomized into class 1 and class 2 based on their position relative to amino acid residue 275. Variant allele frequencies were only available for TCGA and the in-house mCRPC cohorts.

Analysis of whole-genome sequencing data. The bcbio-nextgen pipeline version 1.0.3 was used for the initial steps of tumour whole-genome data analysis. Paired-end reads were aligned to the GRCh38 reference using BWA (bcbio default settings), and structural variant calling was done using LUMPY[50] (bcbio default settings), with the following post-filtering criteria: "(SR> = 1 & PE> = 1 & SU> = 7) & (abs(SVLEN)>5e4) & DP <1000 & FILTER == "PASS". The following settings were chosen to minimize the number of expected germline variants: false discovery rate (FDR) < 0.05 for germline status for both deletions and duplications. Additionally, common structural germline variants were filtered.

Analysis of 10X genomics long-read sequencing data. High-molecular mass DNA from MDA-PCA-2b and LNCaP cell lines was isolated and processed into linked-read next-generation sequencing libraries per the manufacturer's instructions (10X WGS v2 kit). The resulting paired-end sequencing data were sequenced on an Illumina Hi-Seq 2500 instrument and analysed (demultiplexing, alignment, phasing and structural variant calls) using the longranger 2.2.1 pipeline with all default settings. The resulting libraries met all 10X-recommended quality control parameters including molecule size, average phasing length, and sequencing coverage (~50×). Here, we focused on structural variant calls within the *FOXA1* TAD and confirmed the presence of the previously reported *FOXMIND-ETV1* fusions; that is, translocation for MDA-PCA-2b, and balanced insertional translocation for LNCaP. Both cell lines were confirmed to contain three copies of *FOXA1* (that is, one translocated allele and two duplicated alleles).

RNA-seq data pre-processing and primary analysis. RNA-seq data processing—including quality control, read trimming, alignment, and expression quantification by read counting—was carried out as previously described[49], using our standard clinical RNA-seq pipeline CRISP (available at https://github.com/mcieslik-mctp/bootstrap-rnascape). The pipeline was run with default settings for paired-end RNA-seq data of at least 75 bp. The only changes were made for unstranded transcriptome libraries sequenced at the Broad Institute and the TCGA and CCLE cohorts, for which quantification using featureCounts[51] was used in unstranded mode '-s0'. The resulting counts were transformed into fragments per kilobase of transcript per million mapped reads using upper-quartile normalizations as implemented in EdgeR[52]. For mCRPC samples *FOXA1* expression estimates were adjusted by tumour content estimated from whole-exome sequencing (see above) given the highly prostate-specific *FOXA1* expression profile. For the quantification of *FOXMIND* expression levels, a custom approach was necessary given the poor annotation and unspliced nature of this transcript. First, we delineated regions of sense and antisense transcription from the *FOXMIND* ultra-conserved regulatory elements, chr14:37564150-37591250:+ and chr14:37547900-37567150:-, respectively. Next, to make the expression estimates reliable in unstranded libraries, we identified regions of substantial overlap between the sense and antisense RP11-356O9.1 transcripts, and *FOXA1* and *MIPOL1*. These overlaps have been excluded from quantification, resulting in the following trimmed target regions: chr14:37564150-37589500, and chr14:37553500-37567150. Within these regions, the average base-level coverage normalized to sequencing depth was computed as an expression estimate.

Differential expression analyses. All differential expression analyses were done using limma R-package[53], with the default settings for the voom[54], lmFit, eBayes and topTable functions. The contrasts were designed as follows to identify transcriptional signatures of class-1 mutants. Given the mutual exclusivity of the genotypes in primary and metastatic tumours, the overall MCTP mCRPC cohort of 371 cases was partitioned into 4 groups: (1) ETS-fused or *SPOP*-mutant tumours, (2) class-1 mutant tumours, (3) class-2 mutant tumours, and (4) tumours that were wild type for ETS, *SPOP* and *FOXA1*. To avoid confounding effects, the class-2 and ETS and SPOP groups were excluded from class-1 transcriptional analyses. Next, the class-1 samples were contrasted with the wild-type samples with additional independent regressors for assay type (capture vs polyA, as previously described[49], and mutational status (see above) for the following genes and pathways: PI3K, WNT, DRD, *RB1* and *TP53*. In other words, we constructed a design matrix with coefficients for class-1 mutational status, in addition to coefficients for confounding variables and recurrent genetic heterogeneity. This allowed us to estimate the fold changes (expressed logarithmically) and adjusted *P* values associated with *FOXA1* mutations and other genotypes (for example, PI3K status). An analogous procedure was carried out for the primary class-1 samples (TCGA) and for class-2 mutations in mCRPC (MCTP), but given the lack of mutual-exclusivity between class-2 mutations and ETS and SPOP group, only class-1 mutations were excluded.

Pathway and signature enrichment analyses. The Molecular Signatures Database (MSigDB)[55] was used as a source of gene sets comprising cancer hallmarks, molecular pathways, oncogenic signatures and transcription factor targets. The enrichment of signatures was assessed using the parametric random-set method[56], and visualized using the gene-set enrichment analysis (GSEA) enrichment statistic[57] and barcode plots. All *P* values have been adjusted for multiple-hypothesis testing using a false discovery rate correction. To identify putative transcription factors regulating differentially expressed genes, we used the transcription factor prediction tool BART[25]. BART was run with all default settings, and the provided transcription factor databases. We used voom- and limma-based gene-level fold-changes as input to the algorithm.

Detection of structural variants from RNA-seq. The detection of chimeric RNAs (gene fusions, structural variants, circular RNAs and read-through events) was carried out using our previously published[49] in-house toolkit for the comprehensive detection of chimeric RNAs, CODAC (available at https://github.com/mctp/codac). In brief, three separate alignment passes (STAR 2.4.0g1) against the GRCh38 (hg38) reference with known splice junctions provided by Gencode v.27 (ref. [58]) are made for the purposes of expression quantification and fusion discovery. The first pass is a standard paired-end alignment followed by gene-expression quantification. The second and third pass are for the purpose of gene fusion discovery and to enable the chimeric alignment mode of STAR (chimSegmentMin: 10, chimJunctionOverhangMin: 1, alignIntronMax: 150000, chimScoreMin: 1). Fusion detection was carried out using CODAC with default parameters to balance sensitivity and specificity (annotation preset:balanced). CODAC uses MOTR v.2, a custom reference transcriptome based on a subset of Gencode 27 (available with CODAC). Prediction of topology (inversion, duplication, deletion and translocation), and distance (adjacent, breakpoints in two directly adjacent loci; cytoband, breakpoints within the same cytoband based on UCSC genome browser; arm, breakpoints within the same chromosome arm). The high specificity of our pipeline has been assessed through Sanger sequencing[49]. To create fusion circos plots, we have colour-coded the CODAC variants on the basis of the inferred topology of the breakpoints. Unbiased discovery of recurrently rearranged loci has been carried out by breaking the genome into 1.5-Mb windows with a step of 0.5 Mb. For each window, the percentage of patients with at least one RNA break end has been calculated. The resulting genomic windows were ranked and clustered by proximity for visualization. CODAC has the ability to make fusion calls independent of known transcriptome references or annotations and is therefore capable of detecting fusions involving intergenic or poorly annotated regions.

Classification of *FOXA1* locus genomic rearrangements. Structural variants within the *FOXA1* locus have been partitioned into two broad topological patterns: (1) translocations (including inversions and deletions involving distal loci on the same chromosome) and (2) focal duplications. The translocations have been further subdivided into hijacking and swapping events on the basis of their position relative to *FOXMIND* (GRCh38: chr14:37564150-37591250) and *FOXA1*. Hijacking translocations position a translocation partner within the *FOXMIND-FOXA1* regulatory domain (defined as GRCh38: chr14:37547501-37592000, based on manual review of chromatin conformation Hi-C, CTCF, H3K4me1, H3K27ac, evolutionary conservation and synteny data). Swapping

translocations preserve the *FOXMIND-FOXA1* regulatory domain but insert the translocation partner upstream of the *FOXA1* promoter, frequently 'swapping-out' the *TTC6* gene. Notably, one isoform of *TTC6* gene can be transcribed from the bi-directional *FOXA1* promoter. Focal duplications within the *FOXA1* locus have been derived from the CODAC structural-variant output file. In brief, for each case independently, all RNA-seq fusion junctions annotated by CODAC as tandem duplications and overlapping the *FOXA1* topologically associating domain (GRCh38: chr14:37210001-37907919) have been collated and used to infer the minimal duplicated region. Because RNA-seq chimeric junctions generally coincide with splice junctions (limited resolution) and generally cannot be phased (ambiguous haplotype), the inference of minimal duplicated regions makes the necessary and parsimonious assumption that overlapping tandem duplications are due to a single somatic genetic event, and not multiple independent events.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

All raw data for the graphs, immunoblot and gel electrophoresis figures are included in the Source Data or Supplementary Information. All materials are available from the authors upon reasonable request. All the raw next-generation sequencing, ChIP and RNA-seq data generated in this study have been deposited in the Gene Expression Omnibus (GEO) repository at NCBI (accession code GSE123625).

Code availability

All custom data analysis software and bioinformatics algorithms used in this study are publically available on Github: https://github.com/mcieslik-mctp/ and https://github.com/mctp/.

31. Shalem, O. et al. Genome-scale CRISPR–Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
32. Phair, R. D. et al. Global nature of dynamic protein–chromatin interactions in vivo: three-dimensional genome scanning and dynamic interaction networks of chromatin proteins. *Mol. Cell. Biol.* **24**, 6393–6402 (2004).
33. Grimm, J. B. et al. A general method to improve fluorophores for live-cell and single-molecule microscopy. *Nat. Methods* **12**, 244–250 (2015).
34. Pitchiaya, S. et al. Dynamic recruitment of single RNAs to processing bodies depends on RNA functionality. *Mol. Cell* **74**, 521–533 (2019).
35. Swinstead, E. E. et al. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165**, 593–605 (2016).
36. Pitchiaya, S., Androsavich, J. R. & Walter, N. G. Intracellular single molecule microscopy reveals two kinetically distinct pathways for microRNA assembly. *EMBO Rep.* **13**, 709–715 (2012).
37. Shah, N. B. & Duncan, T. M. Bio-layer interferometry for measuring kinetics of protein–protein interactions and allosteric ligand effects. *J. Vis. Exp.* **84**, e51383 (2014).
38. Teng, Y. et al. Evaluating human cancer cell metastasis in zebrafish. *BMC Cancer* **13**, 453 (2013).
39. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
40. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
41. Zhu, L. J. et al. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics* **11**, 237 (2010).
42. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
43. Bailey, T. L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
44. Wilson, S. et al. Developing cancer informatics applications and tools using the NCI genomic data commons API. *Cancer Res.* **77**, e15–e18 (2017).
45. Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
46. Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2**, 401–404 (2012).
47. Wu, Y.-M. et al. Inactivation of *CDK12* delineates a distinct immunogenic class of advanced prostate cancer. *Cell* **173**, 1770–1782 (2018).
48. Cieslik, M. et al. The use of exome capture RNA-seq for highly degraded RNA with application to clinical cancer sequencing. *Genome Res.* **25**, 1372–1381 (2015).
49. Robinson, D. R. et al. Integrative clinical genomics of metastatic cancer. *Nature* **548**, 297–303 (2017).
50. Layer, R. M., Chiang, C., Quinlan, A. R. & Hall, I. M. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* **15**, R84 (2014).
51. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2013).
52. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
53. Smyth, G. K., McCarthy, D. J. & Smyth, G. K. in *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (eds Dudoit, S. & Carey, V. J.) 397–420 (Springer, New York, 2005).
54. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
55. Liberzon, A. et al. The molecular signatures database hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
56. Newton, M. A., Quintana, F. A., Boon, J. A. D., Sengupta, S. & Ahlquist, P. Random-set methods identify distinct aspects of the enrichment signal in gene-set analysis. *Ann. Appl. Stat.* **1**, 85–106 (2007).
57. Subramanian, A., Kuehn, H., Gould, J., Tamayo, P. & Mesirov, J. P. GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* **23**, 3251–3253 (2007).
58. Searle, S. et al. The GENCODE human gene set. *Genome Biol.* **11**, P36 (2010).

Acknowledgements We thank D. Macha, L. Wang, S. Zelenka-Wang, I. Apel, M. Tan, Y. Qiao, A. Delekta, K. Juckette and J. Tien for technical assistance, and S. Gao for assistance with the manuscript. This work was supported by the Prostate Cancer Foundation (PCF), Early Detection Research Network (UO1 CA214170), NCI Prostate SPORE (P50 CA186786) and Stand Up 2 Cancer-PCF Dream Team (SU2C-AACR-DT0712) grants to A.M.C. A.M.C. is an NCI Outstanding Investigator, Howard Hughes Medical Institute Investigator, A. Alfred Taubman Scholar and American Cancer Society Professor. A.P. is supported by a Predoctoral Department of Defense (DoD) - Early Investigator Research Award (W81XWH-17-1-0130). M.C. is supported by a DoD - Idea Development Award (W81XWH-17-1-0224) and a PCF Young Investigator Award.

Author contributions A.P., M.C. and A.M.C. conceived and designed the study; A.P. performed all the experiments with assistance from L.X., T.O., X.W. and S.P. M.C. carried out bioinformatics analyses with assistance from A.P., Y.Z., R.J.L. and P.V. S.-C.C. and A.P. performed zebrafish in vivo experiments. A.P. is responsible for the following experimental figures: Figs. 2b–f, h, 3b–i, 4e, as well as Extended Data Figs. 1a–i, 3b–n, 4a–f, k–n, 5a–k, 6a–l, 7i–o, 8a–h, j, 9a, d, e, 10g. M.C. is responsible for the following computational figures: Figs. 1a–h, 2a, g, 3a, 4a–d, as well as Extended Data Figs. 1j–n, 2a–l, 3a, p, q, 4g–j, o–q, 7a–c, g, h, 9b, c, f–h, 10a–f. Y.Z. is responsible for the following computational figures: Extended Data Figs. 3o, r, s, 7d–f, 8i, k. F.S. and R.W. generated ChIP–seq and RNA-seq libraries. X.C. performed sequencing. F.Y.F. provided genomic validation data. Y.-M.W. and D.R.R. coordinated clinical sequencing. A.P., M.C. and A.M.C. wrote the manuscript and organized the figures.

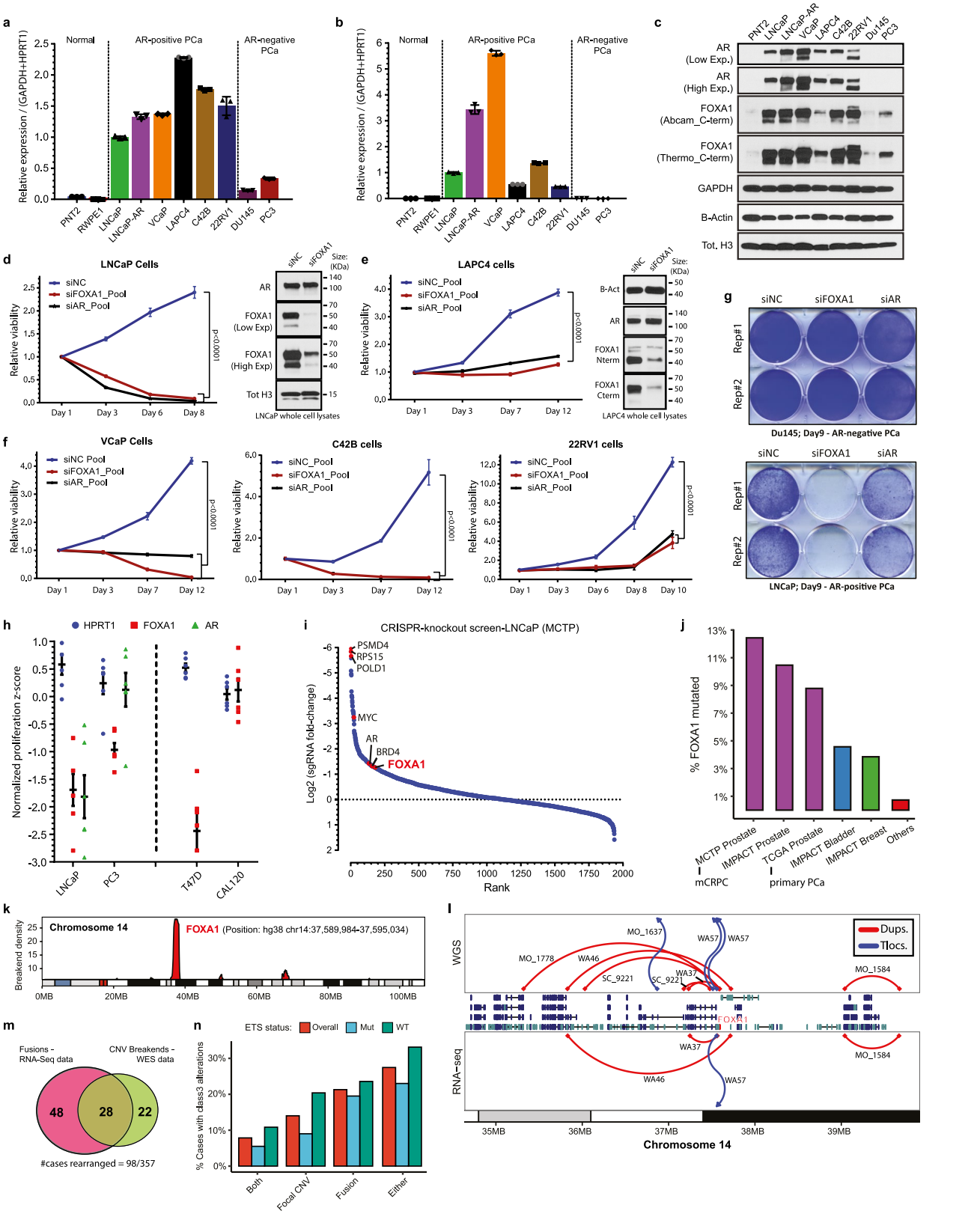
Competing interests The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1347-4.

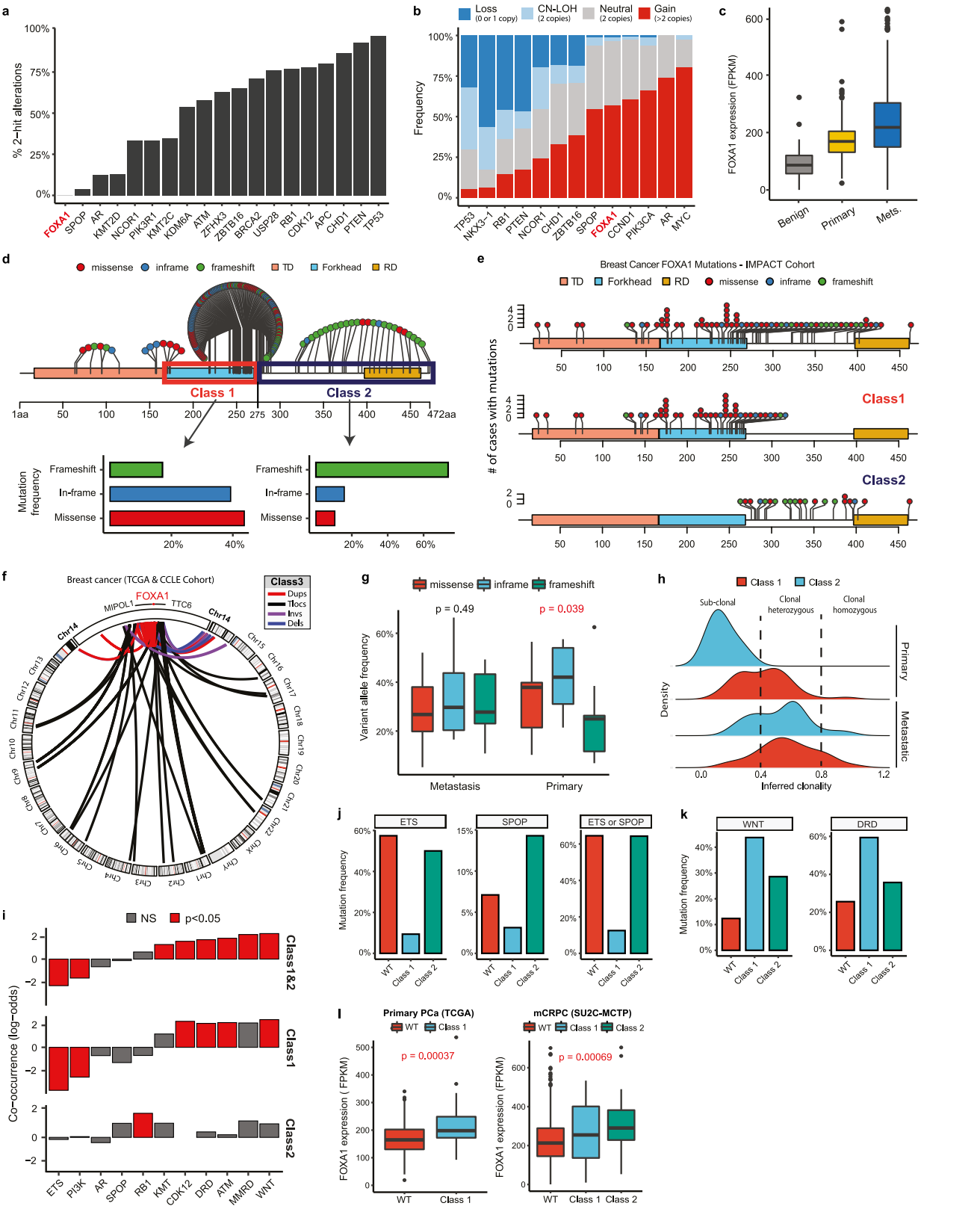
Correspondence and requests for materials should be addressed to A.M.C.

Peer review information *Nature* thanks Myles Brown, William Nelson, Mark A. Rubin and the other anonymous reviewer(s) for their contribution to the peer review of this work.



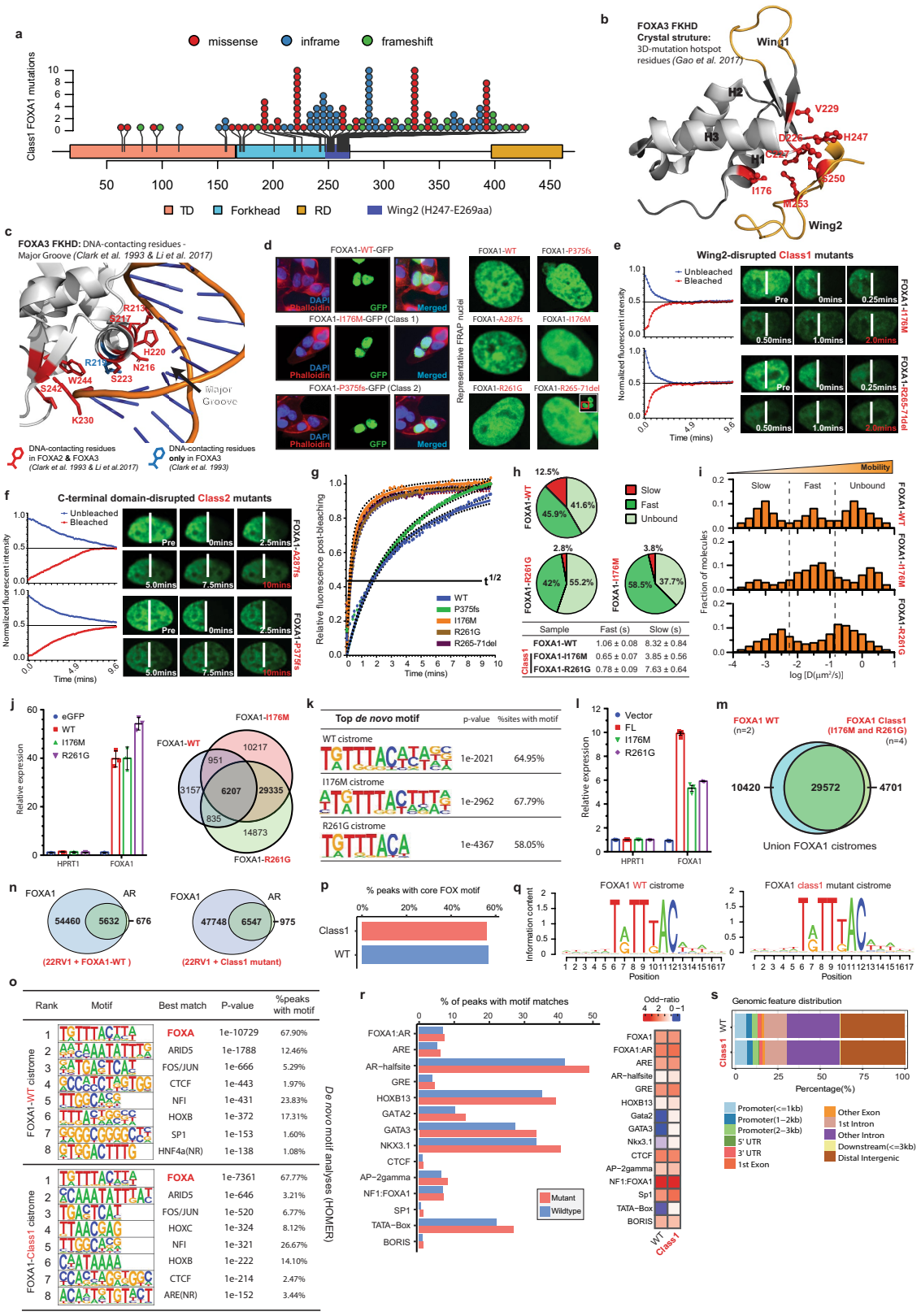
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Functional essentiality and recurrent alterations of FOXA1 in AR$^+$ prostate cancer. a–**c**, AR (**a**) and FOXA1 (**b**) mRNA (qPCR) and (**c**) protein expression in a panel of prostate cancer cells ($n = 3$ technical replicates). Mean ± s.e.m. is shown and dots are individual data points. **d**–**f**, Growth curves of AR$^+$ prostate cancer cells treated with non-targeting control (siNC), *AR*- or *FOXA1*-targeting siRNAs (25 nM at day 0 and 1; $n = 6$ biological replicates). Immunoblots confirm knockdown of FOXA1 protein in LNCaP and LAPC4 72 h after siRNA treatment. For all gel source data, see Supplementary Fig. 1. **g**, Crystal-violet stain of AR$^-$ DU145 prostate cancer and LNCaP (control) cells treated with siNC, *AR*- or *FOXA1*-targeting siRNAs. Results represent 3 independent experiments ($n = 2$ biological replicates). **h**, Averaged proliferation *z*-scores for 6 independent *FOXA1*-targeting sgRNAs extracted from publically available CRISPR Project Achilles data (BROAD Institute) in prostate and breast cancer cells. *HPRT1* and *AR* data serve as negative and positive controls, respectively. Mean ± s.e.m. is shown; dots are proliferative *z*-scores for independent sgRNAs. **i**, Ranked depletion or enrichment of sgRNA read counts from GeCKO-V2 CRISPR knockout screen in LNCaP cells (at day 30) relative to the input sample. Only a subset of genes—including essential controls, chromatin modifiers and transcription factors—is visualized. **j**, Recurrence of *FOXA1* mutations across TCGA, MSK-IMPACT and SU2C cohorts. **k**, Density of break ends (RNA-seq chimeric junctions) within overlapping 1.5-Mb windows along chr14 in mCRPC tumours. **l**, Whole-genome sequencing (WGS) of seven mCRPC index cases with distinct patterns of *FOXA1* translocations (Tlocs) and duplications (Dups), nominated by RNA-seq (WA46, WA37, WA57 and MO_1584) or whole-exome sequencing (MO_1778, SC_9221 and MO_1637). **m**, Concordance of RNA-seq (chimeric junctions) and whole-exome-sequencing-based *FOXA1* locus rearrangements calls (mCRPC cohort). CNV, copy-number variation. **n**, Frequency of *FOXA1* locus rearrangements in mCRPC based on RNA-seq and whole-exome sequencing.



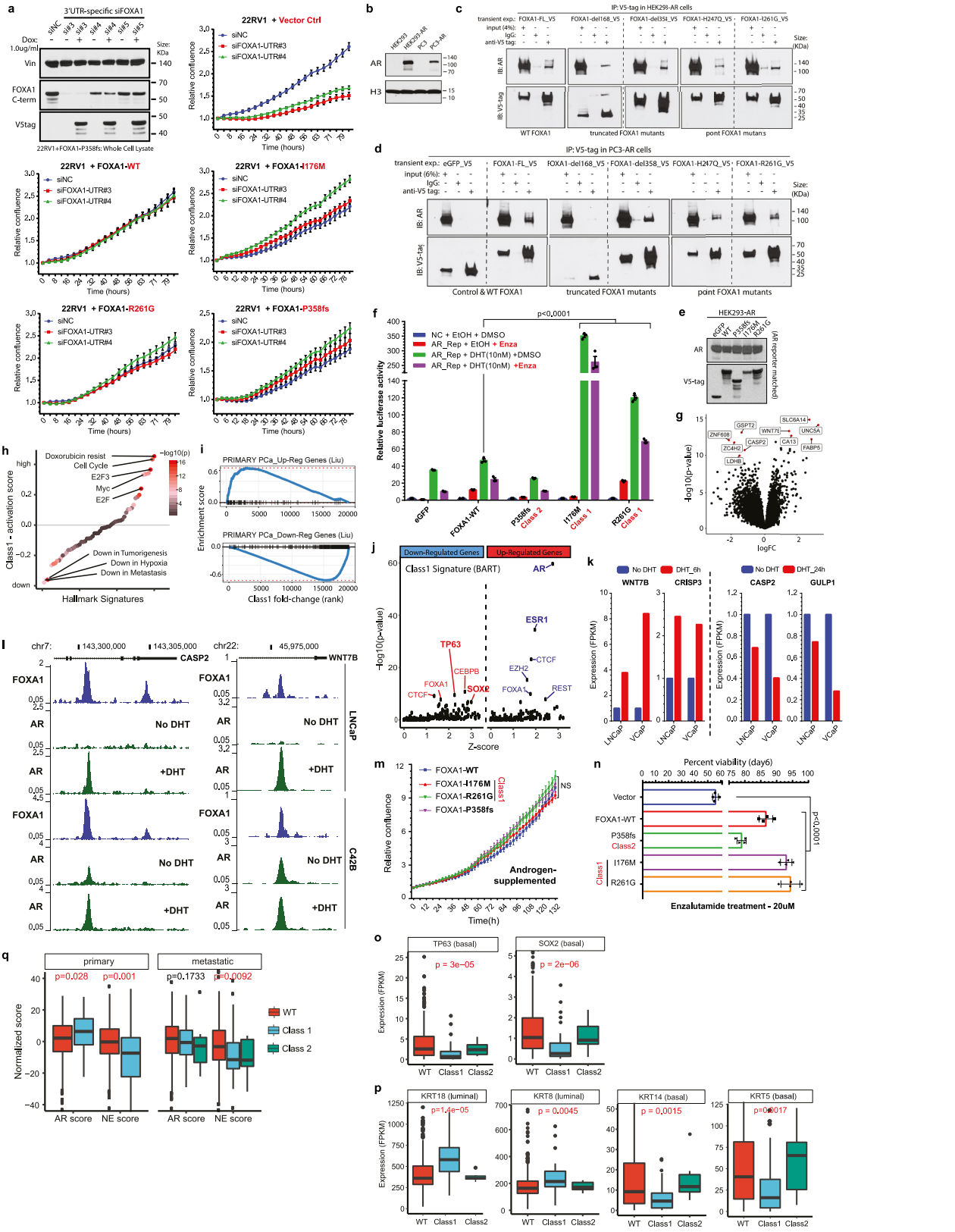
Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Genomic characteristics of the three classes of FOXA1 alterations in prostate and breast cancer. a, **b**, Bi-allelic inactivation (**a**) and copy-number variations (**b**) of *FOXA1* across mCRPC ($n = 371$). CN-LOH, copy-neutral loss of heterozygosity. **c**, FOXA1 expression (RNA-seq) in benign ($n = 51$), primary ($n = 501$) and metastatic ($n = 535$) prostate cancer. **d**, Distribution and functional categorization of *FOXA1* mutations (all cases in the aggregate cohort) on the protein map of FOXA1. **e**, Aggregate and class-specific distribution of *FOXA1* mutations in advanced breast cancer (MSK-IMPACT cohort). **f**, Structural classification of *FOXA1* locus rearrangements in breast cancer (TCGA and CCLE cell lines). **g**, **h**, Variant allele frequency of *FOXA1* mutations by tumour stage (**g**) and clonality estimates of class-1 and class-2 mutations (**h**) in tumour-content-corrected primary prostate cancer ($n = 500$) and mCRPC ($n = 370$) specimens. **i**, Mutual exclusivity or co-occurrence of *FOXA1* mutations (two-sided Fisher's exact test). Mutations in AR, WNT, and PI3K were aggregated at the pathway level. ETS, ETS gene fusions; DRD, DNA repair defects and included alterations in *BRCA1*, *BRCA2*, *ATM* and *CDK12*; MMRD, mismatch repair deficiency (total $n = 371$). **j**, Mutual exclusivity of ETS and/or SPOP ($n = 26$) alterations with FOXA1 ($n = 46$) alterations distinguished by class in mCRPC ($n = 371$). **k**, Co-occurrence of WNT ($n = 58$) and DRD ($n = 107$) pathway alterations with FOXA1 alteration classes in mCRPC ($n = 371$). **l**, Stage- and class-specific increase in FOXA1 expression levels in primary ($n = 500$) and metastatic prostate cancer ($n = 357$). Left, two-sided *t*-test. Right, two-way ANOVA. For all box plots, centre shows median, box marks quartiles 1–3 and whiskers span quartiles 1–3 $\pm$ 1.5 $\times$ IQR.



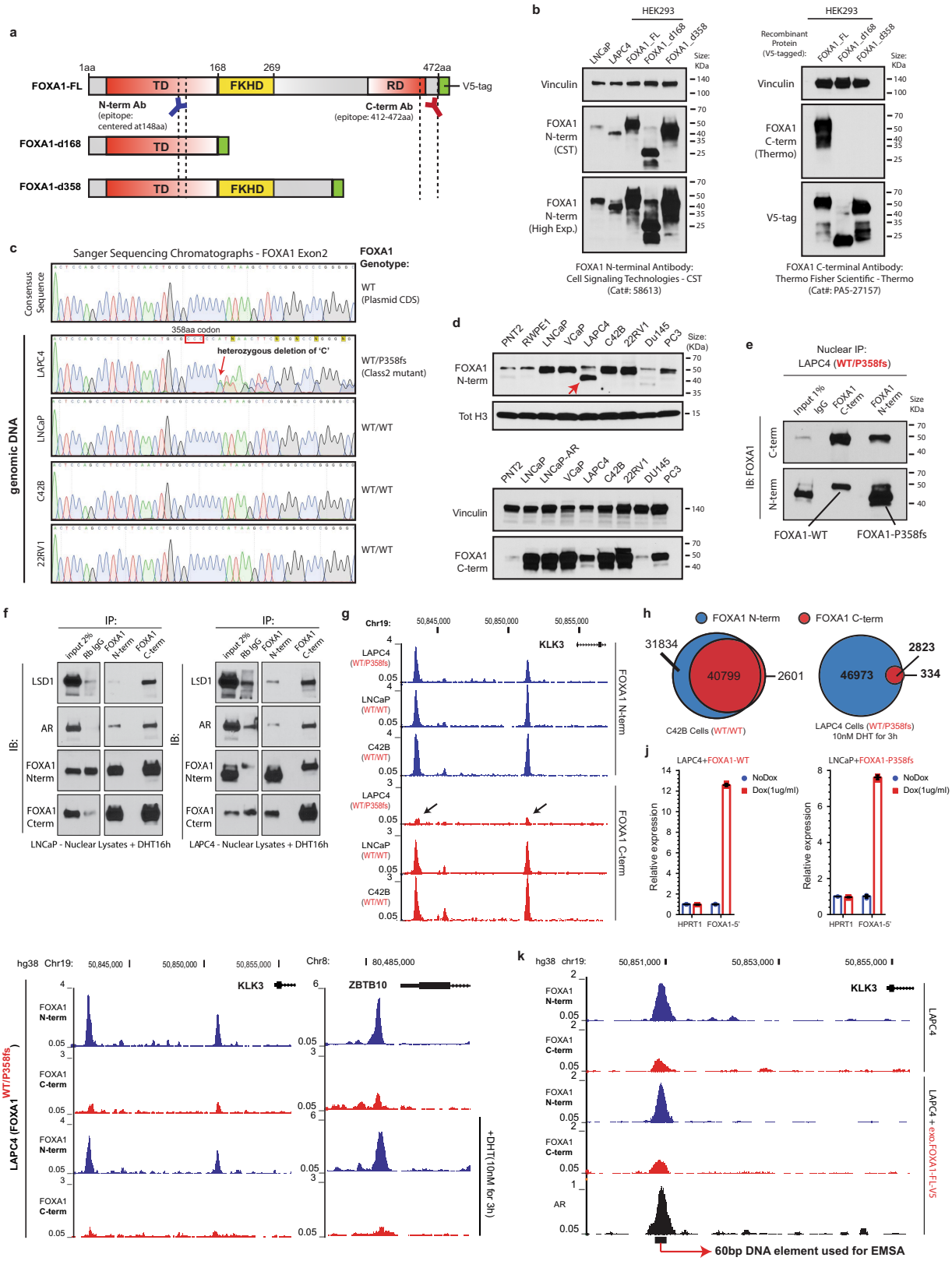
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Biophysical and cistromic characteristics of the class-1 FOXA1 mutants. a, Distribution of class-1 mutations on the protein map of FOXA1. **b**, Three-dimensional structure of FKHD (FOXA3) with visualization of all mutated residues collectively identified as the 3D-mutational hotspot in FOXA1 across cancers. **c**, DNA-bound 3D structure of FKHD with visualization of all residues shown through crystallography to make direct base-specific contacts with the DNA in FOXA2 and FOXA3 proteins. **d**, Representative fluorescent images of nuclei expressing different variants of FOXA1 fused to GFP at the C termini. **e**, **f**, FRAP kinetic plots (left) and representative time-lapse images (right) from pre-bleaching (pre) to 100% recovery (red timestamps) for wing-2-altered class-1 mutants (**e**) and truncated class-2 mutants (that is, A287fs and P375fs) (**f**) ($n = 6$ nuclei per variant; quantified in Fig. 2d). White lines indicate the border between bleached and unbleached areas. **g**, Representative FRAP kinetics in the bleached area for indicated FOXA1 variants. $t_{1/2}$ line indicates the time to 50% recovery. Coloured dots show raw data; superimposed solid curves show a hyperbolic fit with 95% confidence intervals. **h**, Single particle tracking quantification of chromatin-bound (slow and fast) and unbound (freely diffusing) particles of wild-type and class-1 FOXA1 variants, and average chromatin dwell times (mean ± s.d.) for the bound fractions ($n \geq 500$ particles per variant). **i**, Diffusion constant histograms of single particles of wild-type or distinct class-1 FOXA1 mutants. Particles were categorized into chromatin-bound (slow and fast) or unbound fractions using cut-offs marked by dashed lines ($n \geq 500$ particles per variant imaged in 3–5 distinct nuclei). **j**, Left, mRNA expression (qPCR) of labelled FOXA1 variants in stable, isogenic HEK293 cells ($n = 3$ technical replicates). Right, overlaps between FOXA1 wild-type and class-1 mutant cistromes from these cells ($n = 2$ biological replicates). **k**, Top de novo motifs identified from the three FOXA1 cistromes from HEK293 cells (HOMER, hypergeometric test). **l**, mRNA expression (qPCR) of labelled FOXA1 variants in stable, isogenic 22RV1 cells ($n = 3$ technical replicates). For **j** and **l**, centres show mean values and lines mark s.e.m. **m**, Overlap between wild-type ($n = 2$ biological replicates) and class-1 ($n = 4$ biological replicates) cistromes from stable 22RV1 overexpression models. **n**, Overlap between the FOXA1 wild-type and AR union cistromes generated from 22RV1 cells overexpressing wild-type ($n = 2$ biological replicates) or class-1 mutant (I176M or R216G; $n = 2$ biological replicates each) FOXA1 variants. **o**, De novo motif results for the wild-type or class-1 mutant FOXA1-binding sites from prostate cancer cells (HOMER, hypergeometric test). **p**, **q**, Per cent of wild-type or class-1 binding sites with perfect match to the core FOXA1 motif (5′-T[G/A]TT[T/G]AC-3′) (**p**) and the consensus FOXA1 motifs identified from these sites (**q**). **r**, Left, per cent of wild-type or class-1 binding sites containing known motifs of the labelled FOXA1 or AR cofactors. Right, enrichment of the cofactor motifs in the two cistromes relative to the background ($n =$ top 5,000 peaks by score for each variant, see Methods). **s**, Genomic distribution of wild-type and class-1 binding sites in prostate cancer cells.



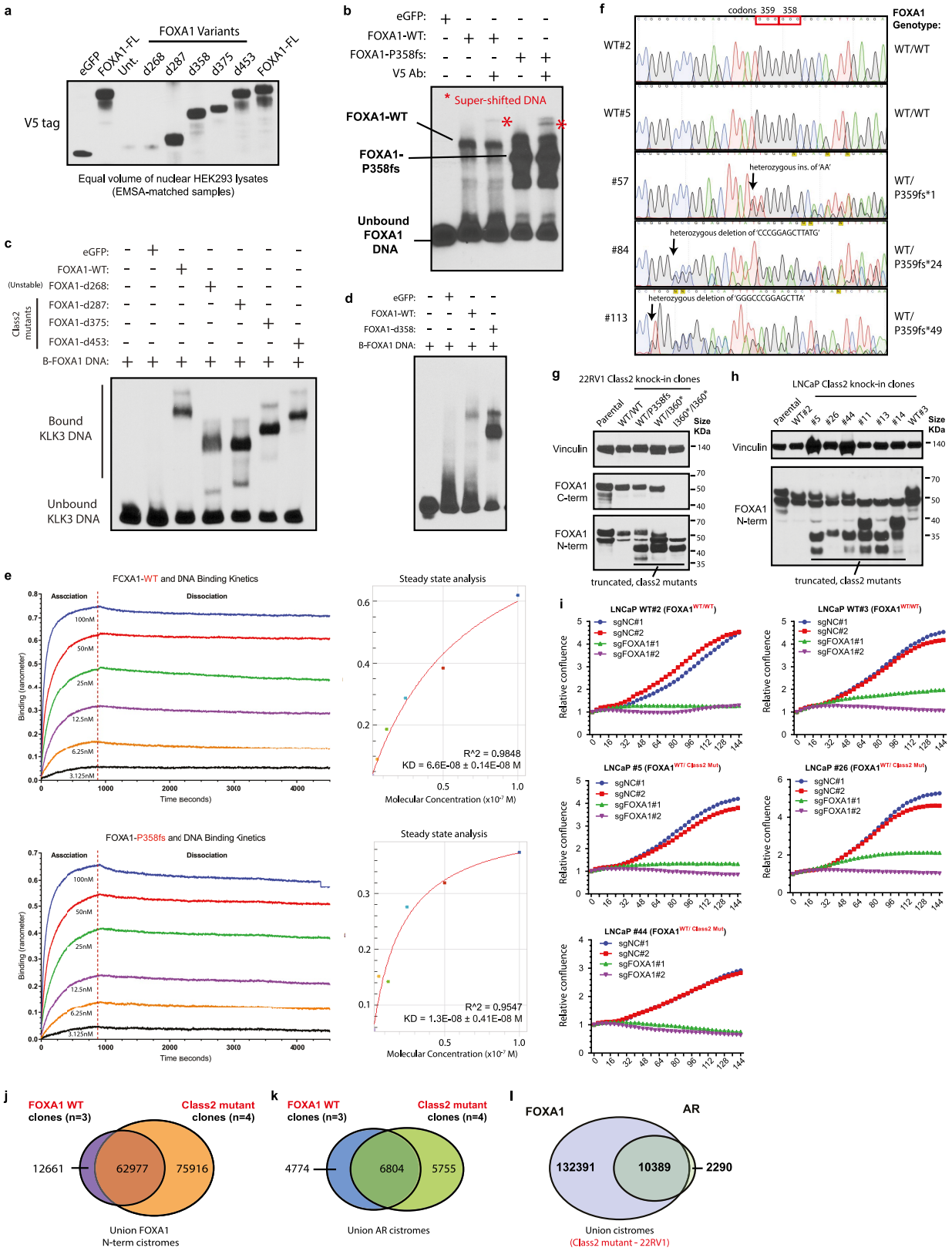
Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | Functional effect of *FOXA1* mutations on oncogenic AR signalling. a, Immunoblot showing expression of endogenous and V5-tagged exogenous FOXA1 proteins in doxycycline (dox)-inducible 22RV1 cells transfected with distinct UTR-specific *FOXA1*-targeting siRNAs (no. 3–5) or a non-targeting control siRNA (siNC). These results represent two independent experiments. IncuCyte growth curves of 22RV1 cells overexpressing empty vector (control), wild-type or mutant FOXA1 variants upon treatment with UTR-specific *FOXA1*-targeting siRNAs ($n = 5$ biological replicates). Mean ± s.e.m. is shown. **b**, Immunoblots confirming stable overexpression of the wild-type AR protein in HEK293 and PC3 cells. **c**, **d**, Co-immunoprecipitation assay of indicated recombinant FOXA1 variants using a V5-tag antibody in HEK293 (**c**) and PC3 (**d**) cells stably overexpressing the AR protein (referred to as HEK293-AR and PC3-AR cells). eGFP is a negative control. FOXA1-FL, full-length wild-type FOXA1. del168 and del358 are truncated FOXA1 variants with only the first 168 amino acids (that is, before the FKHD) or 358 amino acids of the FOXA1 protein. H247Q and R261G are missense class-1 mutant variants. **e**, Immunoblots confirming comparable expression of AR and recombinant FOXA1 variants in AR reporter assay-matched HEK293 lysates. Immunoblots show representative results from 2 or 3 independent experiments and class-1 and class-2 mutants serve as biological replicates. For all gel source data (**a**, **b**–**e**), see Supplementary Fig. 1. **f**, AR dual-luciferase reporter assays with transient overexpression of indicated FOXA1 variants in HEK293-AR cells with or without DHT stimulation and enzalutamide treatment ($n = 3$ biological replicates per group). Mean ± s.e.m. is shown (two-way ANOVA and Tukey's test). **g**, Genes differentially expressed in class-1 tumours from patients ($n = 38$) compared to FOXA1 wild-type tumours (see Methods). The most significant genes are shown in red and labelled (limma two-sided test). **h**, Differential expression of cancer-hallmark signature genes in class-1 mutant prostate-cancer tumours (GSEA statistical test). **i**, Localized, primary prostate cancer gene signature showing concordance between class-1 tumour and primary prostate cancer genes. **j**, BART prediction of specific transcription factors mediating observed transcriptional changes. The significant and strong (z-score) mediators of transcriptional responses in class-1 tumours are labelled (BART, Wilcoxon rank-sum test). **k**, mRNA expression (RNA-seq) of class-1 signature genes in LNCaP and VCaP cells either starved for androgen (no DHT) or stimulated with DHT (10 nM). RNA-seq from two distinct prostate cancer cell lines is shown. **l**, Representative FOXA1 and AR ChIP–seq normalized signal tracks at the *WNT7B* or *CASP2* gene loci in LNCaP and C42B cells. ChIP–seq assays were carried out in two distinct prostate cancer cell lines with similar results. **m**, Growth curves (IncuCyte) of 22RV1 cells overexpressing distinct FOXA1 variants in complete, androgen-supplemented growth medium ($n = 2$ biological replicates). Mean ± s.e.m. is shown. **n**, Per cent viable 22RV1 stable cells, overexpressing either empty vector, wild-type or mutant FOXA1 variants upon treatment with enzalutamide (20 μM for 6 days; $n = 4$ biological replicates). Mean ± s.e.m. is shown. P values in **m** and **n** were calculated using two-way ANOVA and Tukey's test. **o**, **p**, mRNA expression (RNA-seq) of labelled basal and luminal transcription factors or canonical markers in FOXA1 wild-type, class-1 or class-2 mutant tumours in primary prostate cancer (total $n = 500$; two-way ANOVA). **q**, Extent of AR and neuroendocrine (NE) pathway activation in FOXA1 wild-type, class-1 or class-2 mutant cases from both primary ($n = 500$) and metastatic ($n = 370$) prostate cancer. Both AR and NE scores were calculated using established gene signatures (see Methods). Left, two-sided t-test; right, two-way ANOVA. For all box plots, centre shows median, box marks quartiles 1–3 and whiskers span quartiles 1–3 ± 1.5 × IQR.



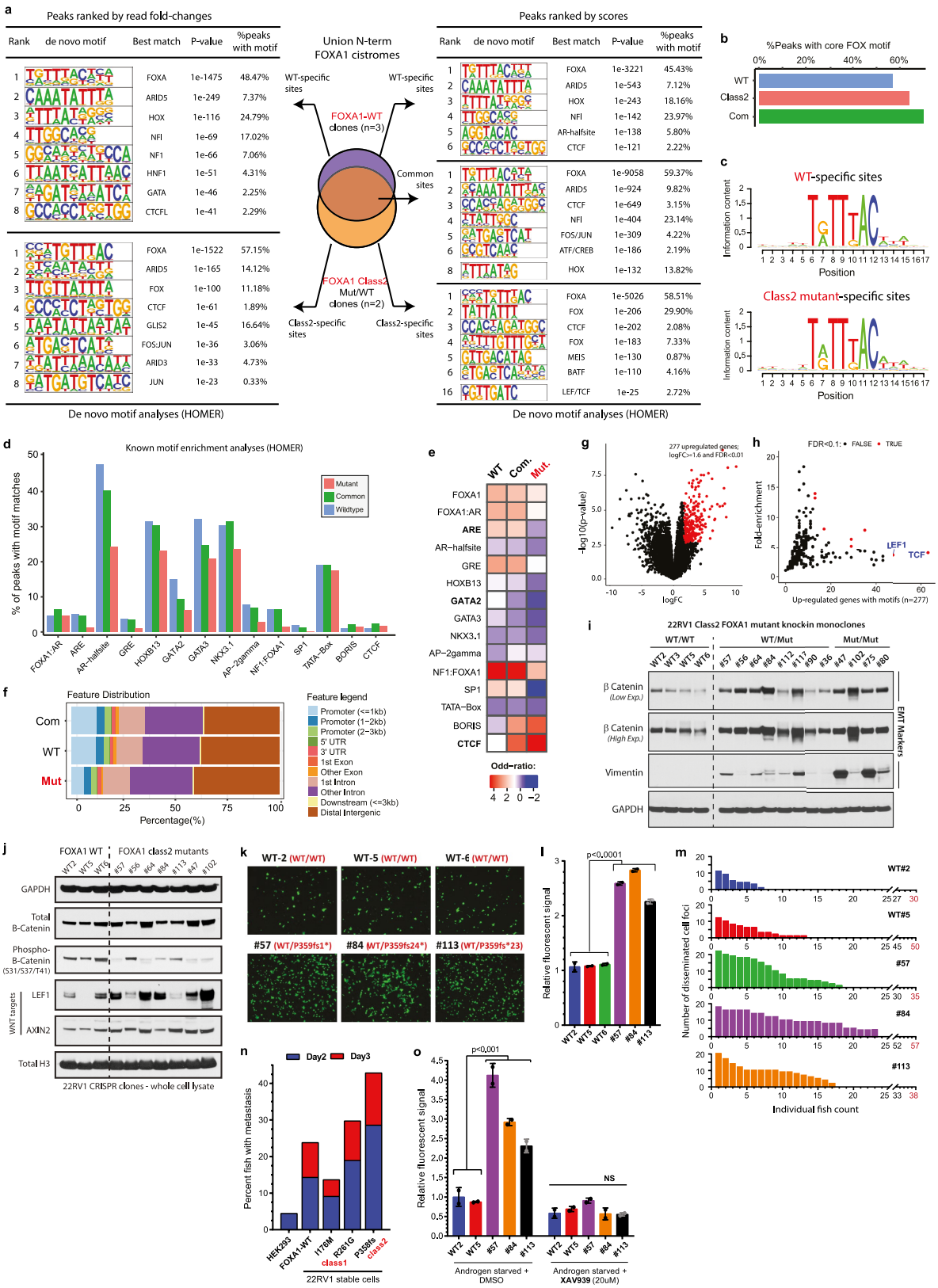
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | DNA-binding dominance of the class-2 FOXA1 mutants. a, FOXA1 protein maps showing the recombinant proteins used to validate the N-terminal (N-term) and C-terminal (C-term) FOXA1 antibodies. **b**, Immunoblots depicting detection of all variants by the N-terminal antibody (left), and of only the full-length wild-type FOXA1 protein by the C-terminal antibody (right). These results were reproducible in two independent experiments. Antibody details are included in the Methods. **c**, Sanger sequencing chromatograms showing the heterozygous class-2 mutation in LAPC4 cells after the P358 codon in exon 2 ($n = 2$ technical replicates). All other tested prostate cancer cell lines were wild type for FOXA1. **d**, Immunoblots confirming the expression of the truncated FOXA1 variant in LAPC4 at the expected approximately 40-kDa size (top, red arrow). The short band is detectable only with the N-terminal (top) FOXA1 antibody and not the C-terminal (bottom) antibody. These results were reproducible in two independent experiments. **e**, Co-immunoprecipitation and immunoblotting of FOXA1 using N-terminal and C-terminal antibodies from LAPC4 nuclei with species-matched IgG used as control. **f**, Nuclear co-immunoprecipitation of FOXA1 from LAPC4 or LNCaP cells stimulated with DHT (10 nM for 16 h) using N-terminal and C-terminal antibodies. Species-matched IgG are controls. Immunoprecipitations and immunoblots in **d**–**f** were reproducible in two and three independent experiments, respectively. For gel source data (**b**, **d**, **e**, **f**), see Supplementary Fig. 1. **g**, FOXA1 N-terminal and C-terminal ChIP–seq normalized signal tracks from FOXA1 wild-type or class-2 mutant prostate cancer cells at canonical AR target *KLK3*. **h**, Left, overlap between global N-terminal and C-terminal FOXA1 cistromes in untreated C42B cells. Right, overlap between global N-terminal and C-terminal FOXA1 cistromes in LAPC4 cells treated with DHT (10 nM for 3 h). **i**, FOXA1 ChIP–seq normalized signal tracks from N-terminal and C-terminal antibodies in LAPC4 cells with or without DHT stimulation (10 nM for 3 h) at *KLK3* and *ZBTB10* loci. ChIP–seq assays in **g** and **i** were carried out in two distinct FOXA1 wild-type prostate cancer cells. For LAPC4 ChIP–seq experiments, results were reproducible in two independent experiments. **j**, mRNA (qPCR) expression of *FOXA1* in LAPC4 cells with exogenous overexpression of wild-type FOXA1 (left), and in LNCaP cells with exogenous overexpression of the P358fs mutant (right) ($n = 3$ technical replicates). Mean ± s.e.m. is shown and dots are individual data values. **k**, FOXA1 ChIP–seq normalized signal tracks from N-terminal and C-terminal antibodies in parental LAPC4 cells and LAPC4 cells overexpressing wild-type FOXA1 at the *KLK3* locus. This experiment was independently repeated twice with similar results. The 60-bp AR- and FOXA1-bound *KLK3* enhancer element used for electrophoretic mobility shift assay (EMSA) is shown.



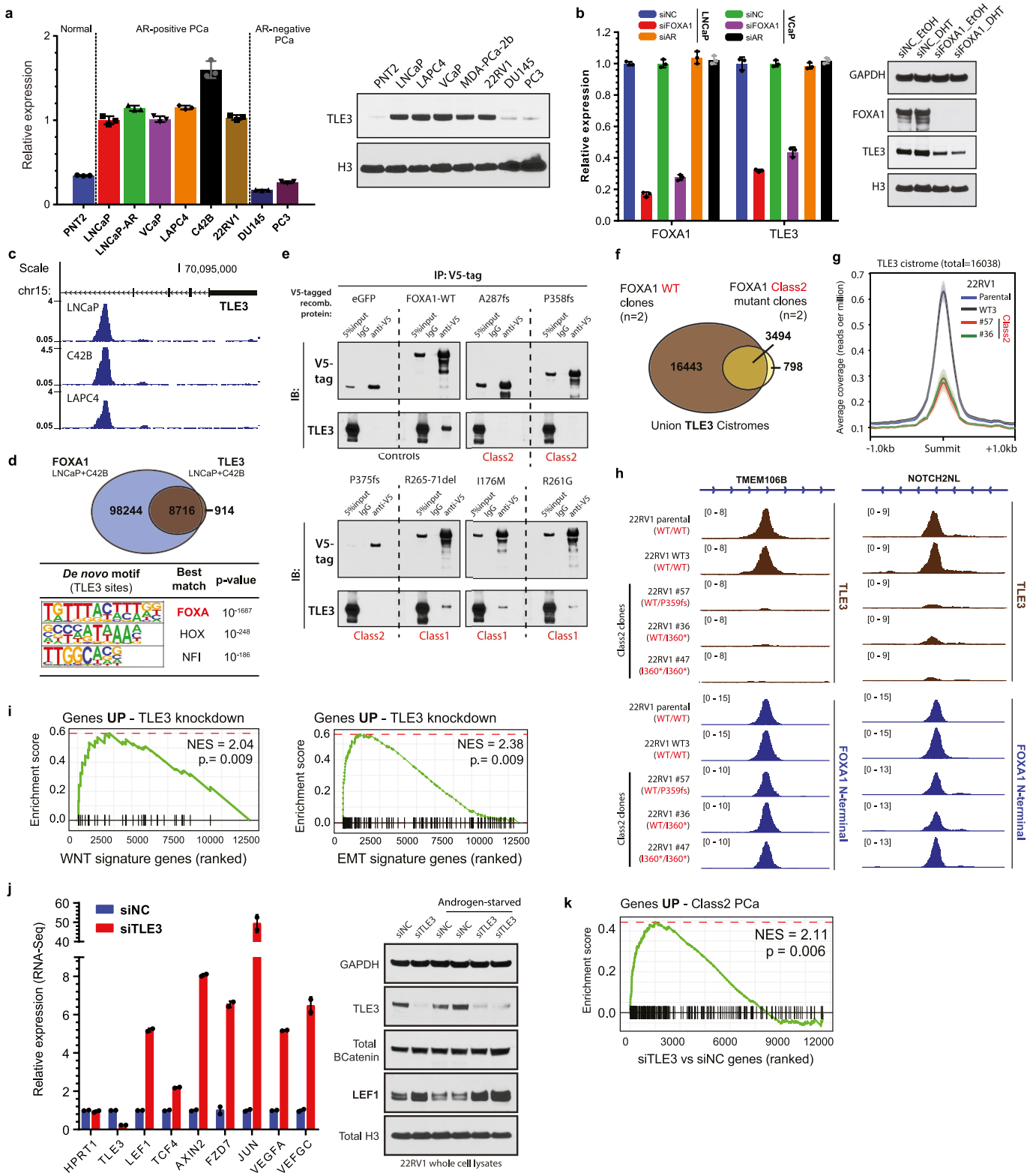
Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | DNA-binding affinity and functional essentiality of the class-2 FOXA1 mutants. a, Immunoblot showing comparable expression of recombinant FOXA1 variants in equal volume of nuclear HEK293 lysates used to perform EMSAs. **b**, Higher exposure of EMSA with recombinant wild-type or P358fs mutant and *KLK3* enhancer element, showing the super-shifted band with addition of the V5 antibody (red asterisks; matched to Fig. 3f). **c**, **d**, EMSA with recombinant wild-type or different class-2 mutants (truncated at 268, 287, 358, 375 and 453 amino acids) and *KLK3* enhancer element. Class-2 mutants display higher affinity than wild-type FOXA1. Each class-2 mutant serves as a biological replicate and these results were reproducible in two independent experiments. **e**, DNA association and dissociation kinetics at varying concentrations of purified wild-type or P358fs class-2 FOXA1 mutants from the biolayer-interferometry assay performed using OctetRED system. Overall binding curves and equilibrium dissociation constants (mean ± s.d.) are shown. These results were reproducible in two independent experiments. **f**, Sanger sequencing chromatograms from a set of 22RV1 CRISPR clones confirming the introduction of distinct indels in the endogenous *FOXA1* allele, resulting in a premature stop codon ($n = 2$ technical replicates). Protein mutations are identified on the right. **g**, Immunoblots showing the expression of endogenous wild-type or class-2 mutant FOXA1 variants in parental and distinct CRISPR-engineered 22RV1 clones. **h**, Immunoblots showing expression of FOXA1 (N-terminal antibody) in parental and CRISPR-engineered LNCaP clones expressing distinct class-2 mutants with truncations closer to the FKHD domain. For gel source data (**a**–**d**, **g**, **h**), see Supplementary Fig. 1. **i**, Growth curves of wild-type or mutant clones upon treatment with the non-targeting or *FOXA1*-targeting sgRNAs and CRISPR–Cas9 protein (see Methods). For **i**, distinct class-2 clones and distinct sgRNAs serve as biological replicates. **j**, **k**, Overlap between union FOXA1 (**j**) and AR (**k**) cistromes from wild-type ($n = 3$ biological replicates) and class-2-mutant ($n = 4$ biological replicates) 22RV1 clones. **l**, Overlap between union FOXA1 and AR cistromes from class-2 mutant 22RV1 cells.



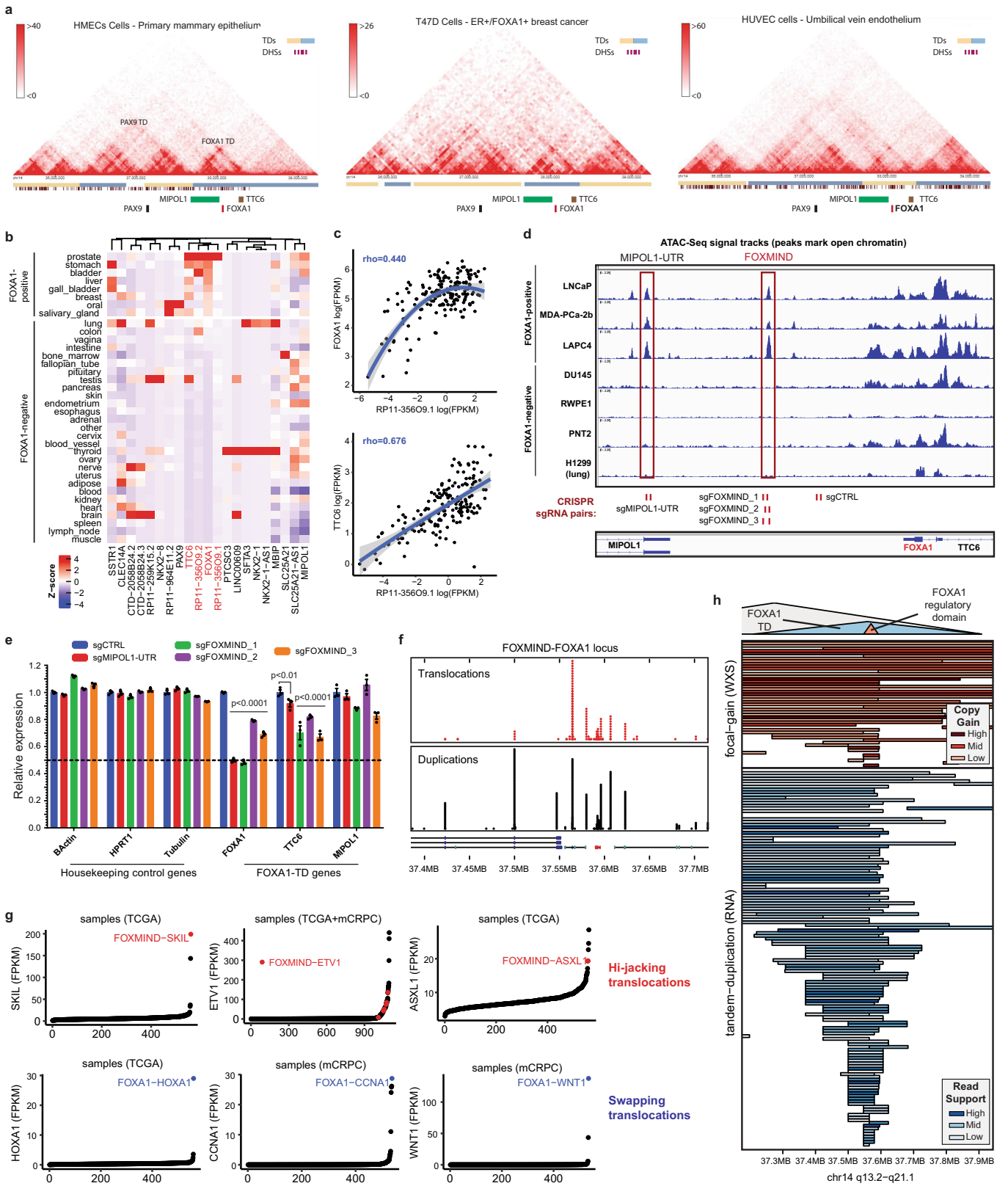
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Cistromic and WNT-driven phenotypic characteristics of the class-2 FOXA1 mutants. a, De novo motif analyses of the wild-type-specific, common and class-2-specific FOXA1-binding site subsets defined from either sequencing-read fold changes (left) or peak-calling scores (right) of ChIP–seq data. Wild-type and class-2 cistromes were generated from $n = 3$ and $n = 2$ independent biological replicates, respectively. Only the top 5,000 or 10,000 peaks from each subset were used as inputs for motif discovery (see Methods) (HOMER, hypergeometric test). **b**, **c**, Per cent of wild-type or class-2 binding sites with perfect match to the core FOXA1 motif (5′-T[G/A]TT[T/G]AC-3′) (**b**) and the consensus FOXA1 motifs identified from these sites (**c**). **d**, **e**, Per cent of binding sites in the three FOXA1-binding-site subsets containing known motifs of the labelled FOXA1 or AR cofactors (**d**), and enrichment of the cofactor motifs in the three binding site subsets relative to the background (**e**). **f**, Genomic distribution of wild-type-specific, common and class-2-specific binding sites in prostate cancer cells. **g**, Differential expression of genes in FOXA1 class-2 mutant CRISPR clones relative to FOXA1 wild-type clones ($n = 2$ biological replicates (limma two-sided test)). **h**, Distinct transcription factor motifs within the promoter (2-kb upstream) of differentially expressed genes. Transcription factors with the highest enrichment (fold change, per cent of upregulated genes with the motif and significance) are highlighted and labelled (two-tailed Fisher's exact test). **i**, Immunoblots showing the expression of β-catenin and vimentin in a panel of wild-type and heterozygous or homozygous class-2 mutant 22RV1 CRISPR clones. **j**, Immunoblots showing the phosphorylation status of β-catenin and expression of direct WNT target genes in select class-2 mutant 22RV1 clones. Immunoblots in **i** and **j** are representative of two independent experiments; every individual clone serves as a biological replicate. For gel source data, see Supplementary Fig. 1. **k**, Representative images of Boyden chambers showing invaded cells stained with calcein AM dye. **l**, Quantified fluorescence signal from invaded cells ($n = 2$ biological replicates per group; two-way ANOVA and Tukey's test). Mean ± s.e.m. is shown and dots are individual data points. **m**, Absolute counts of disseminated cell foci in individual zebrafish embryos as a measure of metastatic burden. **n**, Per cent metastasis at day 2 and day 3 in zebrafish embryos injected with either the normal HEK293 cells (negative controls) or 22RV1 prostate cancer cells virally overexpressing wild-type, class-1 or class-2 mutant FOXA1 variants ($n > 20$ for each group). **o**, Fluorescent signal from the invaded wild-type or class-2-mutant 22RV1 cells after androgen starvation (5% charcoal-stripped serum medium for 72 h) or treatment with the WNT inhibitor XAV939 (20 μM for 24 h; $n = 2$ biological replicates per group; two-way ANOVA and Tukey's test). Mean ± s.e.m. and individual data points are shown.



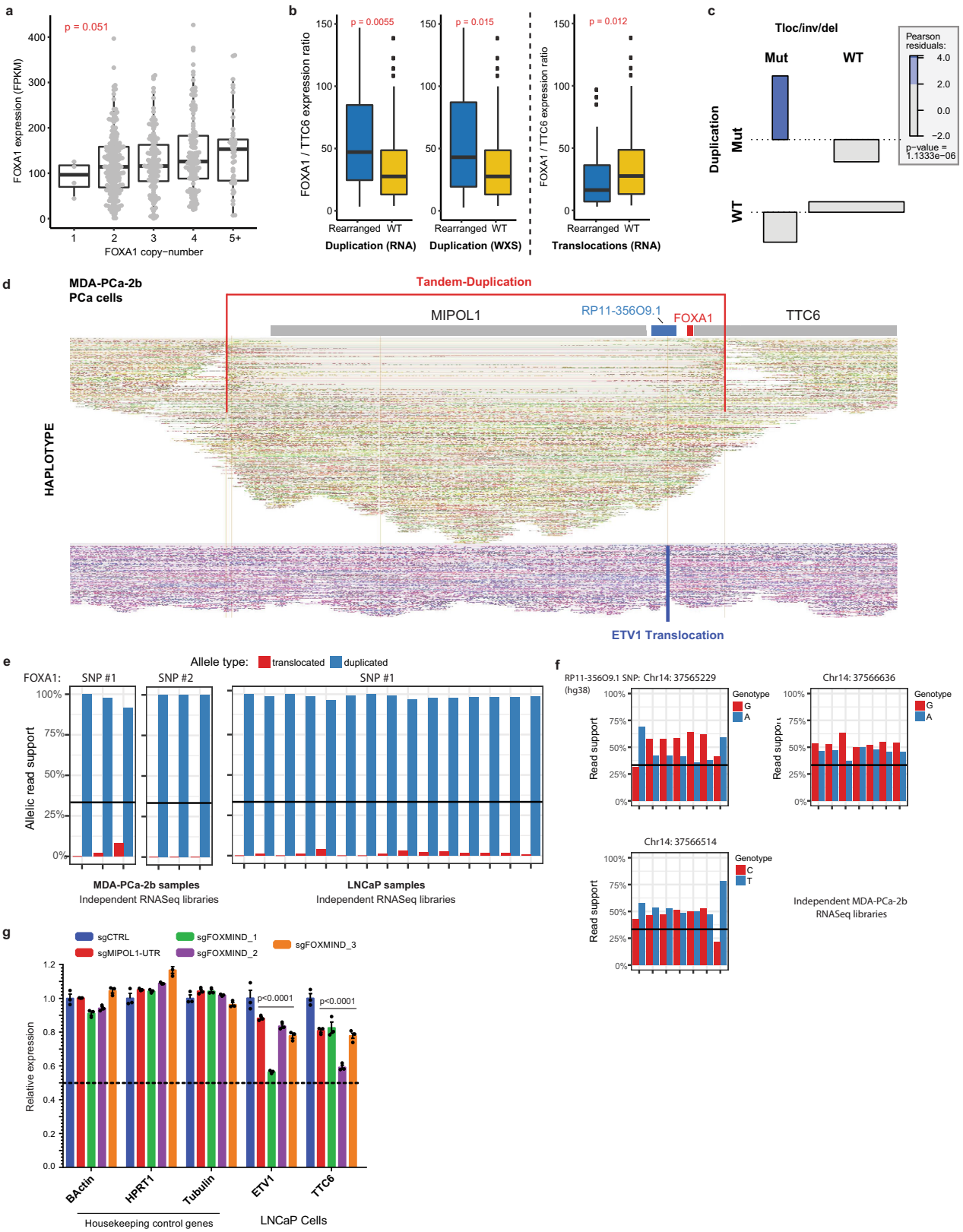
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Functional association of FOXA1 and TLE3 in prostate cancer. a, mRNA (qPCR) and protein (immunoblot) expression of TLE3 in a panel of prostate cancer cells. Mean ± s.e.m. and individual data points are shown. **b**, Left, mRNA expression of *FOXA1* and *TLE3* in LNCaP and VCaP cells treated with siRNAs targeting either *FOXA1* or *AR* ($n = 3$ technical replicates). Two FOXA1 wild-type prostate cancer cells serve as biological replicates. Mean ± s.e.m. and individual data points are shown. Right, protein expression of FOXA1 and TLE3 in matched LNCaP lysates. **c**, FOXA1 N-terminal ChIP–seq normalized signal tracks from LNCaP, C42B and LAPC4 prostate cancer cells at the *TLE3* locus. Each cell line serves as a biological replicate. **d**, Overlap of the union wild-type FOXA1- and TLE3-binding sites from LNCaP and C42B prostate cancer cells ($n = 1$ for each), and top de novo motifs discovered (HOMER, hypergeometric test) in the TLE3 cistrome. **e**, Co-immunoprecipitation assays of labelled recombinant FOXA1 wild-type, class-1 or class-2 variants using a V5-tag antibody in HEK293 cells overexpressing the TLE3 protein. V5-tagged GFP protein was used as a negative control. These results were reproducible in two independent experiments and distinct class-1 and class-2 mutants serve as biological replicates. **f**, Overlap of union TLE3 cistromes from isogenic wild-type ($n = 2$ biological replicates) or heterozygous class-2-mutant ($n = 2$ biological replicates) 22RV1 CRISPR clones. **g**, ChIP peak profile plots from TLE3 ChIP–seq in isogenic FOXA1 wild-type or class-2-mutant 22RV1 clones ($n = 2$ biological replicates each). **h**, Representative TLE3 and FOXA1 ChIP–seq read signal tracks from independent 22RV1 CRISPR clones with or without endogenous *FOXA1* class-2 mutation ($n = 2$ biological replicates each). **i**, GSEA showing significant enrichment of WNT (left) and EMT (right) pathway genes in 22RV1 cells treated with *TLE3*-targeting siRNAs ($n = 2$ biological replicates for each treatment; GSEA enrichment test). **j**, Left, mRNA (RNA-seq) expression of direct WNT target genes in 22RV1 upon siRNA-mediated knockdown of *TLE3* ($n = 2$ biological replicates). Right, Immunoblot showing LEF1 upregulation upon TLE3 knockdown in 22RV1 prostate cancer cells with and without androgen starvation (representative of two independent experiments). For gel source data (**a**, **b**, **e**, **j**), see Supplementary Fig. 1. **k**, Gene enrichment plots showing significant enrichment of class-2 upregulated genes upon *TLE3* knockdown in 22RV1 cells ($n = 2$ biological replicates for each treatment; GSEA enrichment test).



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Topological, physical and transcriptional characteristics of the *FOXA1* locus in normal tissues and prostate cancer. a, HI-C data (from: http://promoter.bx.psu.edu/hi-c/view.php) depicting conserved topological domains within the *PAX9* and *FOXA1* syntenic block in normal and FOXA1$^+$ cancer cell lines. DHSs, DNase I hypersensitive sites. **b**, Highly tissue-specific patterns of gene expression within the *PAX9* and *FOXA1* syntenic block. Tissues were dichotomized into FOXA1$^+$ and FOXA1$^-$ on the basis of FOXA1 expression levels; genes were subject to unsupervised clustering. *z*-score normalization was performed for each gene across all tissues. **c**, Correlation of RP11-356O9.1 (Methods) and *FOXA1* or *TTC6* expression levels across metastatic tissues ($n = 370$; Spearman's rank correlation coefficient). The 95% confidence interval is shown. **d**, Representative ATAC-seq ($n = 1$) read signal tracks from normal basal epithelial prostate (RWPE1 and PNT2 cells) or prostate cancer cells. Cells are grouped on the basis of expression of FOXA1, and differentially pioneered loci are marked with red boxes. CRISPR sgRNA pairs used for genomic deletion of the labelled elements are shown at the bottom. Distinct FOXA1$^+$ and FOXA1$^-$ cell lines serve as biological replicates for ATAC-seq. **e**, mRNA (qPCR) expression of housekeeping control genes, genes located within the *FOXA1* topologically associated domain, and *MIPOL1* in VCaP cells treated with CRISPR sgRNA pairs targeting a control site (sgCTRL), *FOXMIND* or the *MIPOL1* UTR regulatory element (see Extended Data Fig. 2c for sgRNA binding sites). Distinct sgRNA pairs cutting at *FOXMIND* serve as biological replicates. Mean ± s.e.m. is shown ($n = 3$ technical replicates; two-way ANOVA and Tukey's test). **f**, Distribution of tandem duplication and translocation break ends (chimeric junctions or copy-number segment boundaries) focused at the *FOXMIND-FOXA1* regulatory domain. **g**, Outlier expression of genes involved in translocations with the *FOXA1* locus. Translocations positioning a gene between *FOXMIND* and *FOXA1* (hijacking) are shown on top (red). Translocations positioning a gene upstream of the *FOXA1* promoter (swapping) are shown on the bottom (blue). **h**, Inferred duplications within the *FOXA1* locus on the basis of RNA-seq (tandem break ends) and whole-exome sequencing (copy-gains), zoomed-in at the *FOXA1* topologically associating domain.



Extended Data Fig. 10 | See next page for caption.

Extended Data Fig. 10 | Transcriptional and genomic characteristics of class-3 FOXA1 rearrangements in prostate cancer. a, Dosage sensitivity of the *FOXA1* gene. Expression of *FOXA1* (RNA-seq) across mCRPC tumours ($n = 370$) as a function of gene ploidy (as determined by absolute copy number at the *FOXA1* locus (two-way ANOVA)). **b**, Relative expression of *FOXA1* (within the minimally amplified region) to *TTC6* (outside the amplified region) in rearranged ($n = 50$) (duplication or translocation) versus wild-type ($n = 320$) *FOXA1* loci (two-sided *t*-test). For all box plots, centre shows median, box marks quartiles 1–3 and whiskers span quartiles 1–3 ± 1.5 × IQR. **c**, Association plot visualizing the relative enrichment of cases with both translocation and duplications within the *FOXA1* locus ($n = 370$). Overabundance of cases with both events is quantified using Pearson residuals. Significance of this association is based on the χ^2 test without continuity correction. Inv, inversion; del, deletion. **d**, *FOXA1* locus visualization of linked-read (10X platform) whole-genome sequencing of the MDA-PCA-2b cell line. Alignments on the haplotype-resolved genome are shown in green and purple. Translocation and tandem-duplication calls are indicated in blue and red, respectively. **e**, Monoallelic expression of *FOXA1* cell lines with *FOXMIND-ETV1* translocations in MDA-PCA-2b ($n = 6$ biological replicates) and LNCaP ($n = 15$ biological replicates). Phasing of *FOXA1* SNPs to structural variants is based on linked-read sequencing (Methods). **f**, Biallelic expression of the RP11-356O9.1 transcript assessed using three distinct SNPs in MDA-PCA-2b cells that contain *ETV1* translocation into the *FOXA1* locus ($n = 7$ biological replicates). **g**, mRNA (qPCR) expression of *ETV1* and *TTC6* upon sgRNA-mediated disruption of the *FOXMIND* or the *MIPOL1* UTR enhancer in LNCaP cells, which also contain *ETV1* translocation into the *FOXA1* locus (see Extended Data Fig. 9d for sgRNA binding sites). Distinct sgRNA pairs cutting at *FOXMIND* serve as biological replicates. Mean ± s.e.m. are shown ($n = 3$ technical replicates; two-way ANOVA and Tukey's test).

LETTER

https://doi.org/10.1038/s41586-019-1369-y

scSLAM–seq reveals core features of transcription dynamics in single cells

Florian Erhard[1,6]*, Marisa A. P. Baptista[1,6], Tobias Krammer[2,6], Thomas Hennig[1], Marius Lange[3,4], Panagiota Arampatzi[5], Christopher S. Jürges[1], Fabian J. Theis[3,4], Antoine-Emmanuel Saliba[2]* & Lars Dölken[1,2]*

Single-cell RNA sequencing (scRNA-seq) has highlighted the important role of intercellular heterogeneity in phenotype variability in both health and disease[1]. However, current scRNA-seq approaches provide only a snapshot of gene expression and convey little information on the true temporal dynamics and stochastic nature of transcription. A further key limitation of scRNA-seq analysis is that the RNA profile of each individual cell can be analysed only once. Here we introduce single-cell, thiol-(SH)-linked alkylation of RNA for metabolic labelling sequencing (scSLAM-seq), which integrates metabolic RNA labelling[2], biochemical nucleoside conversion[3] and scRNA-seq to record transcriptional activity directly by differentiating between new and old RNA for thousands of genes per single cell. We use scSLAM-seq to study the onset of infection with lytic cytomegalovirus in single mouse fibroblasts. The cell-cycle state and dose of infection deduced from old RNA enable dose–response analysis based on new RNA. scSLAM-seq thereby both visualizes and explains differences in transcriptional activity at the single-cell level. Furthermore, it depicts 'on–off' switches and transcriptional burst kinetics in host gene expression with extensive gene-specific differences that correlate with promoter-intrinsic features (TBP–TATA-box interactions and DNA methylation). Thus, gene-specific, and not cell-specific, features explain the heterogeneity in transcriptomes between individual cells and the transcriptional response to perturbations.

SLAM-seq[3] involves briefly exposing cells to the nucleoside analogue 4-thiouridine (4sU). 4sU is incorporated into new RNA during transcription and converted to a cytosine analogue using iodoacetamide (IAA) before RNA sequencing. Sequencing reads originating from new RNA can be identified within the pool of total RNA reads on the basis of characteristic U- to-C conversions. We applied the SLAM-seq technique to resolve the onset of lytic mouse cytomegalovirus (CMV) infection at the single-cell level. After optimization for single-cell sequencing (scSLAM-seq) (Fig. 1, Supplementary Methods), we performed scSLAM-seq on 107 single mouse fibroblast cells and in parallel analysed global transcriptional changes of matched, larger (1×10^5) populations of cells ($n = 2$) using (bulk) SLAM-seq. After quality filtering for cells with more than 2,500 detectable genes (Extended Data Fig. 1a), the remaining samples (49 CMV-infected, 45 uninfected cells) displayed all the characteristics of high-quality scSLAM-seq libraries (Extended Data Fig. 1b), including U-to-C conversion rates of between 4% and 6% (Extended Data Fig. 1c, d). Incorporation of 4sU is thus both efficient and uniform at the single-cell level.

Owing to rates of 4sU incorporation of about 1 in 50–200 nucleotides, up to 50% of all SLAM-seq reads that originate from new RNA may not contain U-to-C conversions. To overcome this problem, we developed 'globally refined analysis of newly transcribed RNA and decay rates using SLAM-seq' (GRAND-SLAM)—a Bayesian method to compute the ratio of new to total RNA (NTR) in a fully quantitative manner including credible intervals[4] (Fig. 1). Here we report GRAND-SLAM 2.0 for the parallel analysis of hundreds of SLAM-seq libraries derived from single cells. The accuracy of quantification is further improved by analysing long reads (150 nucleotides) in paired-end mode (see Supplementary Methods), which allows 4sU conversions to be reliably distinguished from sequencing errors within the overlapping sequences (Extended Data Fig. 1c, d). We obtained accurate measurements (90% credible interval < 0.2) for thousands of genes per cell, thereby approaching the overall sensitivity of scRNA-seq (Extended Data Fig. 1e) and achieving high correlation ($R > 0.73$) with bulk SLAM-seq (Extended Data Fig. 1f).

Unbiased principal component analysis (PCA) of highly variable cellular genes (see Supplementary Methods, Extended Data Fig. 1g) could not separate CMV-infected from uninfected cells for either total RNA or old RNA, and only slightly for new RNA (Fig. 2a, Extended Data Fig. 1h–j). Intercellular heterogeneity thus exceeded the virus-induced changes, which are hardly detectable in total RNA by two hours post-infection (h.p.i.), owing to the slow turnover of mammalian mRNAs (see Extended Data Fig. 2a–d, Supplementary Methods and Supplementary Table 1). By contrast, PCA on the NTR separated uninfected from infected cells with high precision (Fig. 2a) and demonstrated a clear positive correlation with the extent of viral gene expression (Pearson's correlation coefficient $R = 0.59$, $P = 7.3 \times 10^{-6}$) (Extended Data Fig. 1j).

Recent findings reported that intronic reads from scRNA-seq data can be used to estimate time derivatives of gene expression in individual cells termed 'RNA velocities'[5]. These indicate the future trajectory of individual cells in low-dimensional projections of gene-expression space. However, infected cells could not be separated from uninfected cells by an unbiased PCA computed on the respective RNA velocities, or on the expression profiles projected into the future using the velocities, or directly on intron/exon ratios (Extended Data Fig. 3a). To compare scSLAM-seq directly with RNA velocities computed for a larger population of cells, we performed 10x Genomics Chromium droplet-based scRNA-seq on hundreds of uninfected ($n = 793$) and CMV-infected ($n = 353$) cells using the same experimental conditions. Although PCA on mature transcripts (exonic reads only) did not separate uninfected and infected cells, the distinction was possible using intron/exon ratios (Extended Data Fig. 3b). However, no meaningful directionalities in the RNA velocities of both the scSLAM-seq and 10x data were observed (Fig. 2b). We used new and total RNA levels obtained by scSLAM-seq to replace intronic and exonic read levels and determine 'NTR velocities'. Notably, these further discriminated infected from non-infected cells (Fig. 2c).

To compare NTRs and RNA velocity directly, we asked which of them could best predict whether a gene was upregulated or downregulated in large cell populations. Although this was possible to some extent using RNA velocities computed from dozens or hundreds of cells in scSLAM-seq or 10x data, respectively (area under receiver operating characteristic curve (AUC) values of 0.68 and 0.74), they were

[1]Institute for Virology and Immunobiology, Julius-Maximilians-University Würzburg, Würzburg, Germany. [2]Helmholtz Institute for RNA-based Infection Research (HIRI), Helmholtz-Center for Infection Research (HZI), Würzburg, Germany. [3]Institute of Computational Biology, Helmholtz Zentrum München–German Research Center for Environmental Health, Neuherberg, Germany. [4]Department of Mathematics, Technische Universität München, Munich, Germany. [5]Core Unit Systems Medicine, University of Würzburg, Würzburg, Germany. [6]These authors contributed equally: Florian Erhard, Marisa A. P. Baptista, Tobias Krammer. *e-mail: florian.erhard@uni-wuerzburg.de; emmanuel.saliba@helmholtz-hiri.de; lars.doelken@uni-wuerzburg.de

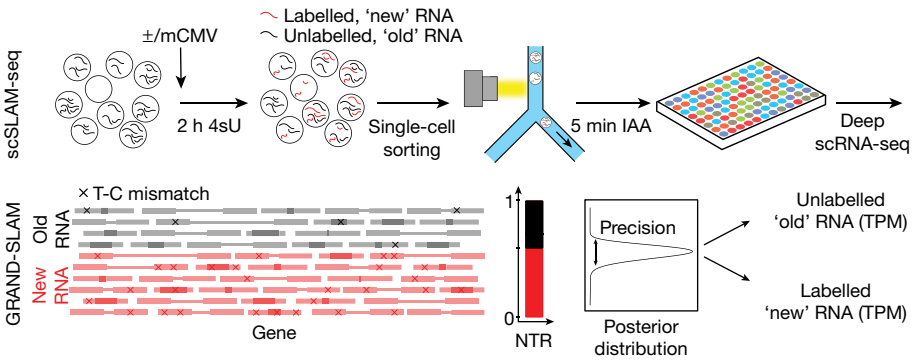


Fig. 1 | scSLAM-seq resolves transcriptional activity at the single-cell level. Overview of scSLAM-seq (top) and GRAND-SLAM (bottom) approaches. Top, nascent transcripts are labelled before or after CMV infection by adding 500 μM 4sU to the cell culture medium for 2 h. After single-cell sorting and RNA isolation, 4sU is converted into a cytosine analogue by IAA and SMART-seq libraries are prepared and sequenced. Bottom, GRAND-SLAM identifies thymine-to-cytosine mismatches and estimates both the NTR and the expression of old and new RNA. TPM, transcript per millions.

outperformed by using NTRs (AUC > 0.94) (Fig. 2d, Extended Data Fig. 3c). Moreover, NTRs could be determined for substantially more regulated genes ($n = 583$) than velocities ($n = 131$ and $n = 295$, from scSLAM-seq and 10x data, respectively). NTRs and RNA velocities thus convey different—but complementary—information (Extended Data Fig. 3d), and can be merged into NTR velocities to predict the future state of a cell more reliably.

A fundamental question in virology is why the infection of one cell results in extensive lytic virus replication whereas the infection of a second cell does not. At 2 h.p.i., CMV infection in most cells has already progressed from the 'immediate-early' (restricted to a few genes) to the 'early' phase of infection, in which most viral genes are already transcribed. Although most viral transcripts in all cells were new, we also observed substantial amounts of old RNAs of some viral genes (Extended Data Fig. 4). This represents virion-associated RNA that is delivered to the infected cell by the incoming virus particles[6]. It correlated well with virion-associated RNA isolated from the virus stock source ($R = 0.48$; Extended Data Fig. 5a, b) and thus provides

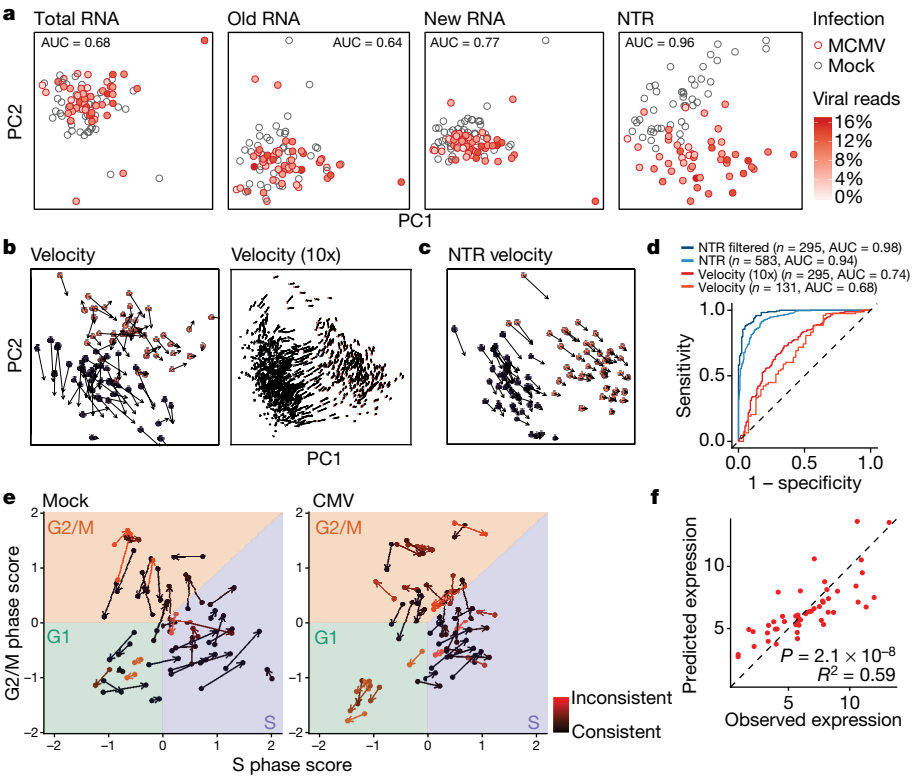


Fig. 2 | scSLAM-seq and NTR velocities. a, PCA of highly variable cellular genes separates infected ($n = 2$ replicates, 49 cells) and uninfected ($n = 2$ replicates, 45 cells) cells based only on the NTR and not on total, old or new RNA. MCMV, mouse cytomegalovirus. **b**, PCA computed on regulated genes (see Extended Data Fig. 2c) with RNA velocities indicated by arrows for the scSLAM-seq (left; uninfected, $n = 2$ replicates, 43 cells; infected, $n = 2$ replicates, 44 cells) or the 10x (right; uninfected, $n = 2$ replicates, 793 cells; infected, $n = 2$ replicates, 353 cells) data. **c**, PCA generated as in **b** (left) but showing NTR velocities computed using new and total, instead of unspliced and spliced, counts. **d**, NTRs or RNA velocities computed on the scSLAM-seq or 10x data were used to predict upregulation or downregulation of genes regulated more than twofold in bulk RNA-seq analysis. The number of used genes and AUC values are indicated. NTRs were also filtered to the same set of genes as the 10x data (same cells as in **b** for scSLAM-seq and 10x). **e**, Cell-cycle progression (G1, S and G2/M phases) for the uninfected (mock, $n = 2$ replicates, 45 cells) and CMV-infected ($n = 2$ replicates, 49 cells) cells, showing trajectories based on cell-cycle states deduced from old RNA (base of arrows) and total RNA (tip of arrows). **f**, Scatter plot comparing the predicted extent of viral gene expression per individual cell ($n = 2$ replicates, 49 cells) on the basis of cell-cycle state and dose of infection with the observed expression. The coefficient of determination (R^2) is indicated. *P* value determined by likelihood ratio test.

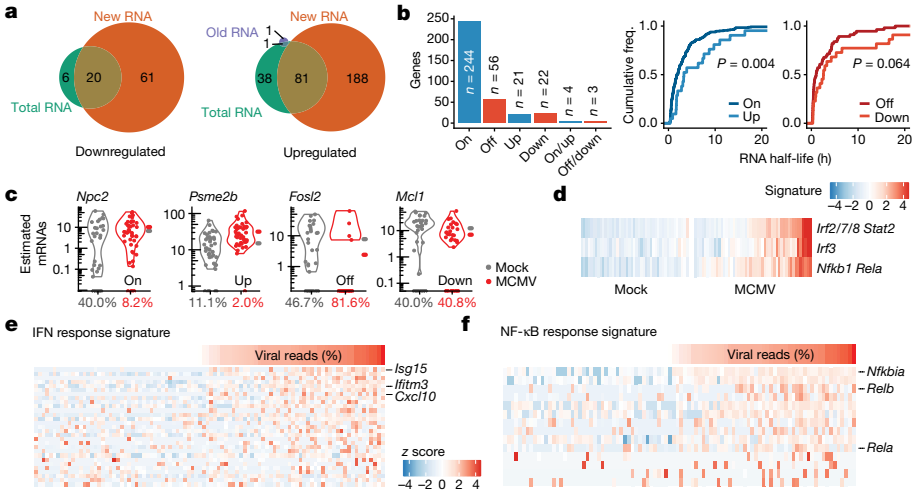


Fig. 3 | scSLAM-seq depicts the mode of gene regulation and differentially activated pathways in single cells. a, Venn diagrams representing upregulated and downregulated genes after CMV infection called using SC2P (10% false discovery rate) and on the basis of total, new or old RNA. **b**, Left, regulated genes in new RNA are classified by SC2P according to the mode of regulation (on/up and off/down). Middle and right, RNA half-life distributions of up and down genes are compared to on and off genes, respectively. *P* values determined by two-sided Wilcoxon rank-sum test. **c**, *Npc2*, *Psme2b*, *Fosl2* and *Mcl1* illustrate the different modes of regulation in new RNA. Grey denotes uninfected cells ($n = 2$ replicates, 45 cells); red denotes CMV-infected cells ($n = 2$ replicates, 49 cells). The percentage of cells not expressing a transcript is indicated below each violin plot. Average expression levels are indicated to the right of each plot. The *y* axis indicates read counts normalized to spike-in controls. Violins show kernel density estimates. **d**, Unbiased pathway analysis (PAGODA)[12] of new RNA revealed transcriptional signatures and transcription factor targets associated with uninfected and CMV-infected cells. **e**, **f**, Standardized expression levels of IFN-responsive (**e**) and NF-κB-responsive (**f**) genes of each cell are depicted for new RNA. Cells are ordered according to the total percentage of viral reads.

a surrogate marker for the dose of infection that each individual cell has received. By contrast, new viral RNA reflects the infection efficacy. The dose of infection explained 52% of the variance in infection efficacy (Extended Data Fig. 5c). Accordingly, the amount of old viral RNA in the cells that hardly expressed any new viral RNA (Extended Data Fig. 4) was much lower than in the other cells, which indicates that these cells were not less permissive to CMV infection but instead received a much lower dose of virus. On the basis of cell-cycle signature genes, the cell-cycle state both at the beginning and at the end of 4sU metabolic labelling can be inferred from old and total RNA, respectively, thereby providing cell-cycle trajectories[7] (Fig. 2e). Although lytic infection was initiated at all cell-cycle stages, cells infected during G1 phase resulted in significantly stronger viral gene expression and cell-cycle disruption ($P < 0.05$; Extended Data Fig. 5d, e). This increased the explained variance in new viral gene expression to 59% (Fig. 2f). The efficiency by which lytic viral gene expression is initiated at the single-cell level in fibroblasts is thus well explained by the interaction of the dose of infection and cell cycle.

To assess the effects of CMV infection on cellular gene expression, we identified differentially expressed genes between CMV-infected and uninfected single cells in total, new and old RNA using single-cell two-phase testing procedure (SC2P)[8] (Supplementary Table 2). Most (more than 60%) of the downregulated (61 out of 87) and upregulated (188 out of 309) genes could be uncovered only by specifically considering new RNA (Fig. 3a). Bimodality of gene expression is a well-described feature in single cells[8,9]. A bimodally expressed gene is undetectable in a subpopulation of cells but expressed in others. scSLAM-seq directly visualizes whether the promoter of a given gene in a cell was 'on' within the studied time frame. Notably, we found that most CMV-induced changes in new RNA are much more consistent with on–off dynamics than with upregulation or downregulation (Fig. 3b, c).

CMV infection induces a strong type I interferon (IFN) and NF-κB response during the first two hours of infection[10,11]. Gene set analysis[12] from predicted transcription factor targets[13] and Gene Ontology terms demonstrated that the activation of both IFN and NF-κB was highly virus-dose dependent. However, although IFN activation was limited to about half of the infected cells, NF-κB activation occurred in most CMV-infected cells (Fig. 3d, Extended Data Fig. 5f, Supplementary Tables 3, 4). Virus dose-dependent activation of NF-κB in all cells is consistent with M45 tegument protein-mediated activation of NF-κB at or upstream of the IKK kinase complex[11]. By contrast, IFN responses require the detection of pathogen-associated molecular patterns[14], are subject to autocrine and paracrine signalling, and may thus show greater variability between individual cells[15]. To perform a more-focused analysis, we defined both NF-κB- and IFN-responsive gene sets specific to CMV infection on the basis of previously published data on NF-κB induction[16] and IFN treatment[2] using our bulk SLAM-seq data (Supplementary Table 5). The magnitudes of the IFN (Fig. 3e) and NF-κB (Fig. 3f) responses varied markedly between individual cells but both were highly correlated with viral gene expression (Spearman's $\rho > 0.52$, $P < 3 \times 10^{-4}$) (Extended Data Fig. 5g, h). Most NF-κB- and IFN-inducible gene expression thus arises in the most-strongly infected cells, with the weakest responses being induced in cells infected during S phase (Extended Data Fig. 5i, j).

Transcriptional activity at the single-cell level is not a continual process but consists of intermittent bursts of transcription separated by minutes to hours of transcriptional inactivity, indicative of temporarily non-permissive promoters[17,18]. We reasoned that such bursting kinetics of a gene should be detectable by scSLAM-seq. To assess bursting kinetics globally, we defined gene-wise burst scores (*B*-scores) as the standard deviation of the NTR distribution from all uninfected cells in which RNA of the respective gene could be reliably quantified (90% credible interval of the NTR < 0.2; Supplementary Table 6; $n = 5{,}540$). *B*-scores obtained from two independent biological replicates were highly correlated ($R = 0.74$) (Extended Data Fig. 6a). In some cases, extreme *B*-scores close to 0.5 corresponded to genes with only a single or very few mRNA molecules (either new or old) detected in each cell (Extended Data Fig. 6b).

Current scRNA-seq protocols either provide unique molecular identifiers (UMIs) to estimate the number of captured mRNA molecules and are strand-specific but only clone transcript ends (for example, 10x Genomics Chromium scRNA-seq), or provide full-length mRNAs but no UMIs and lose strand-specificity (for example, Smart-seq2[19]). Our Smart-seq-based scSLAM-seq approach—at least in part—encompasses all three features. The random incorporation of 4sU

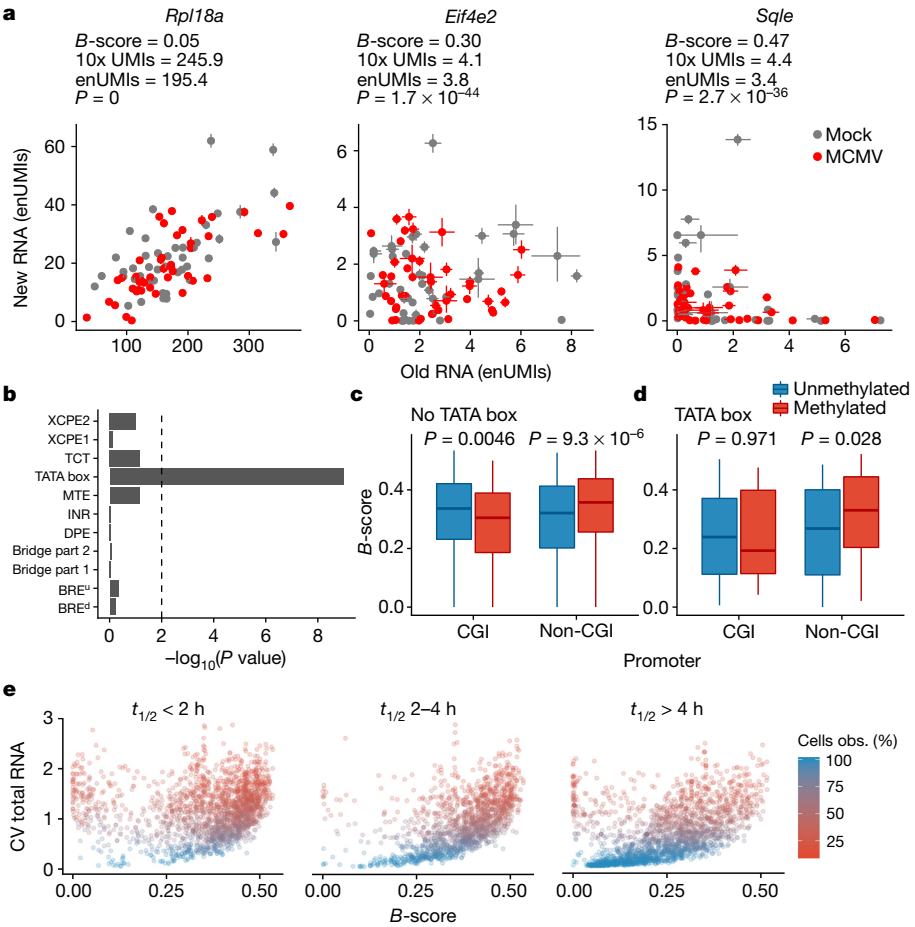


Fig. 4 | scSLAM-seq reveals bursting kinetics and core features of heterogeneity in transcription. a, Representative genes with different burst scores (*B*-scores; increasing from left to right). The estimated numbers of captured molecules based on enUMIs are shown, with associated 90% credible intervals. The average enUMI count, the UMI count from the 10x experiment, and the *P* value from a χ^2 test are indicated. Grey denotes uninfected cells ($n = 2$ replicates, 45 cells); red denotes CMV-infected cells ($n = 2$ replicates, 49 cells). **b**, Promoter structure analysis reveal TATA boxes to be highly enriched in promoters of genes with low *B*-scores. The log-transformed *P* values are indicated (two-sided Wilcoxon rank-sum test, $n = 2$ replicates, 45 cells). BRE^u and BRE^d, TFIIB recognition elements upstream and downstream of the TATA box, respectively; DPE, downstream promoter element; INR, initiator element; MTE, motif 10 element; TCT, polypyrimidine initiator element; XCPE1 and XCPE2, X core promoter element 1 and 2, respectively. **c**, *B*-score distributions for CpG island (CGI) and non-CpG island (non-CGI) promoters stratified by DNA methylation status in bisulfite sequencing experiments. Only promoters without TATA boxes were considered. *P* values for differences of scores in strata are indicated (two-sided Wilcoxon rank-sum test, $n = 2$ replicates, 45 cells). Box plots denote the median (centre line) and interquartile range (box), with whiskers extending to three times the interquartile range. **d**, As in **c** but considering only TATA-box promoters. **e**, *B*-scores for each gene are scattered against the coefficient of variation (CV) of total RNA across cells. Genes were stratified according to their RNA half-life ($t_{1/2}$). The percentage of non-drop-out cells is indicated.

into mRNA molecules provides nucleotide conversion-based unique molecular identifiers (nUMIs), which enabled us to estimate a lower bound of the number of new mRNA molecules sampled per cell and gene, and—by extrapolation (enUMIs)—also of old mRNAs (Extended Data Figs. 6c–h, 7). On the basis of these conservative estimates of the number of sampled mRNAs, we found that more than 30% (1,718 out of 5,540; adjusted $P < 0.01$, χ^2 test) of all detectable genes had greater variance (*B*-score) than expected from sampling (see Fig. 4a for examples). There was only a negligible correlation of the *B*-score with expression levels (Extended Data Fig. 8a). Moreover, the observed heterogeneity in the NTR did not result from cell-cycle-dependent differences (Extended Data Fig. 8b–e) but was associated with mRNA half-life (Extended Data Fig. 8f).

Unbiased Gene Ontology overrepresentation analysis revealed high gene *B*-scores to be associated with functional categories such as protein phosphorylation and ubiquitination (Supplementary Table 7). Promoter analyses identified six motifs that were significantly enriched for either low (TATA box motif) or high (CG-rich and purine-rich motifs) (Extended Data Fig. 9; Supplementary Table 8) *B*-scores. Correctly placed TATA boxes were most highly enriched ($P < 10^{-8}$), consistent with promoters of the TATA box driving frequent transcriptional bursts on a timescale of minutes[20]. No association with other core promoter motifs was observed (Fig. 4b). The CG-rich motifs could either correspond to binding sites of specific transcription factors that exhibit an oscillatory activation pattern or reflect CpG-rich regions within the respective promoters. More than 50% of mammalian transcription initiates from promoters close to CpG islands (CGI promoters), which represent CpG-rich regions of dozens to hundreds of nucleotides. Hypermethylation of CpG islands is an epigenetic control mechanism of gene silencing[21]. Bisulfite sequencing data from mouse fibroblasts[22] revealed significant correlation of methylated CGI promoters with low *B*-scores, whereas methylated non-CGI promoters tended to exhibit high *B*-scores (Fig. 4c). The same was also observed for TATA-box-containing promoters (Fig. 4d). We confirmed these results by repeating the analyses for the 1,718 genes with significant *B*-scores (Extended Data Fig. 10a) and on the top 50% most-strongly expressed genes (Extended Data Fig. 10b). Although we cannot fully exclude that allelic differences in transcriptional activity and karyotype complexity of the cell line used here account for some of the observed effects[23], this does not explain the strong correlation of *B*-scores to promoter-intrinsic features. We propose a model in which DNA methylation within gene promoters is involved in the transient

regulation of promoter activity on the timescale of hours by temporarily rendering gene promoters non-permissive. Finally, *B*-scores were highly correlated with the observed heterogeneity in the transcriptomes of individual cells (Fig. 4e). Gene promoter-specific features thus represent a major contributor to intercellular heterogeneity.

Metabolic labelling using 4sU is applicable to all major model organisms including vertebrates, insects, plants and yeast. The purine analogue 6-thioguanine (6sG) now also enables G-to-A conversions by oxidative-nucleophilic-aromatic substitution (TimeLapse-seq chemistry)[24]. Short, consecutive pulses of 4sU and 6sG followed by thiol-(SH)-mediated nucleoside conversions may enable two independent recordings of transcriptional activity in single cells. Finally, scSLAM-seq combined with CRISPR-based perturbations will greatly improve the sensitivity of the respective approaches to decipher the molecular mechanisms with major implications for developmental biology, infection and cancer.

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The sequencing data and gene tables are available from the Gene Expression Omnibus (GEO) with accession number GSE115612. The script files are available at zenodo (doi: 10.5281/zenodo.1299119). GRAND-SLAM is available for non-commercial use at http://software.erhard-lab.de.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests are available at https://doi.org/10.1038/s41586-019-1369-y.

Received: 7 August 2018; Accepted: 10 June 2019;
Published online 10 July 2019.

1. Wagner, A., Regev, A. & Yosef, N. Revealing the vectors of cellular identity with single-cell genomics. *Nat. Biotechnol.* **34**, 1145–1160 (2016).
2. Dölken, L. et al. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14**, 1959–1972 (2008).
3. Herzog, V. A. et al. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* **14**, 1198–1204 (2017).
4. Jürges, C., Dölken, L. & Erhard, F. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* **34**, i218–i226 (2018).
5. La Manno, G. et al. RNA velocity of single cells. *Nature* **560**, 494–498 (2018).
6. Terhune, S. S., Schröer, J. & Shenk, T. RNAs are packaged into human cytomegalovirus virions in proportion to their intracellular concentration. *J. Virol.* **78**, 10390–10398 (2004).
7. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
8. Wu, Z., Zhang, Y., Stitzel, M. L. & Wu, H. Two-phase differential expression analysis for single cell RNA-seq. *Bioinformatics* **34**, 3340–3348 (2018).
9. Shalek, A. K. et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* **498**, 236–240 (2013).
10. Marcinowski, L. et al. Real-time transcriptional profiling of cellular and viral gene expression during lytic cytomegalovirus infection. *PLoS Pathog.* **8**, e1002908 (2012).
11. Krause, E., de Graaf, M., Fliss, P. M., Dölken, L. & Brune, W. Murine cytomegalovirus virion-associated protein M45 mediates rapid NF-κB activation after infection. *J. Virol.* **88**, 9963–9975 (2014).
12. Fan, J. et al. Characterizing transcriptional heterogeneity through pathway and gene set overdispersion analysis. *Nat. Methods* **13**, 241–244 (2016).
13. Pachkov, M., Balwierz, P. J., Arnold, P., Ozonov, E. & van Nimwegen, E. SwissRegulon, a database of genome-wide annotations of regulatory sites: recent updates. *Nucleic Acids Res.* **41**, D214–D220 (2013).
14. Lio, C.-W. J. et al. cGAS-STING signaling regulates initial innate control of cytomegalovirus infection. *J. Virol.* **90**, 7789–7797 (2016).
15. Rand, U. et al. Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Mol. Syst. Biol.* **8**, 584 (2012).
16. Hinata, K., Gervin, A. M., Jennifer Zhang, Y. & Khavari, P. A. Divergent gene regulation and growth effects by NF-κ B in epithelial and mesenchymal cells of human skin. *Oncogene* **22**, 1955–1964 (2003).
17. Hodges, C., Bintu, L., Lubkowska, L., Kashlev, M. & Bustamante, C. Nucleosomal fluctuations govern the transcription dynamics of RNA polymerase II. *Science* **325**, 626–628 (2009).
18. Tantale, K. et al. A single-molecule view of transcription reveals convoys of RNA polymerases and multi-scale bursting. *Nat. Commun.* **7**, 12248 (2016).
19. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* **10**, 1096–1098 (2013).
20. Zoller, B., Nicolas, D., Molina, N. & Naef, F. Structure of silent transcription intervals and noise characteristics of mammalian genes. *Mol. Syst. Biol.* **11**, 823 (2015).
21. Koch, A. et al. Analysis of DNA methylation in cancer: location revisited. *Nat. Rev. Clin. Oncol.* **15**, 459–466 (2018).
22. Maza, I. et al. Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33**, 769–774 (2015).
23. Reinius, B. & Sandberg, R. Random monoallelic expression of autosomal genes: stochastic transcription and allele-level regulation. *Nat. Rev. Genet.* **16**, 653–664 (2015).
24. Kiefer, L., Schofield, J. A. & Simon, M. D. Expanding the nucleoside recoding toolkit: revealing RNA population dynamics with 6-thioguanosine. *J. Am. Chem. Soc.* **140**, 14567–14570 (2018).

Acknowledgements We thank J. Vogel and S. Gorski for comments on the manuscript and T. Walzthoeni (Bioinformatics Core Facility, Institute of Computational Biology,Helmholtz Zentrum München) for bioinformatics support. The work was funded by the European Research Council (ERC-2016-CoG 721016-HERPES and ERC-2018-PoC-DL3 832409 T-GRAND-SLAM) and Infect-ERA grant eDEVILLI (031L0005B) to L.D. The Helmholtz Institute for RNA-based Infection Research (HIRI) supported this work with a seed grant through funds from the Bavarian Ministry of Economic Affairs and Media, Energy and Technology (grant allocation no. 0703/68674/5/2017 and 0703/89374/3/2017). F.J.T. and M.L. acknowledge financial support by the Graduate School QBM (GSC 1006); F.J.T. was supported by BMBF grants 01IS18036A and 01IS18053A, the German Research Foundation (DFG) within the Collaborative Research Centre 1243, Subproject A17, the Helmholtz Association (Incubator grant sparse2big, ZT-I-0007) and the Chan Zuckerberg Initiative DAF (advised fund of Silicon Valley Community Foundation, 182835). M.L. and P.A. acknowledge financial support from the Joachim Herz Stiftung and IZKF, respectively.

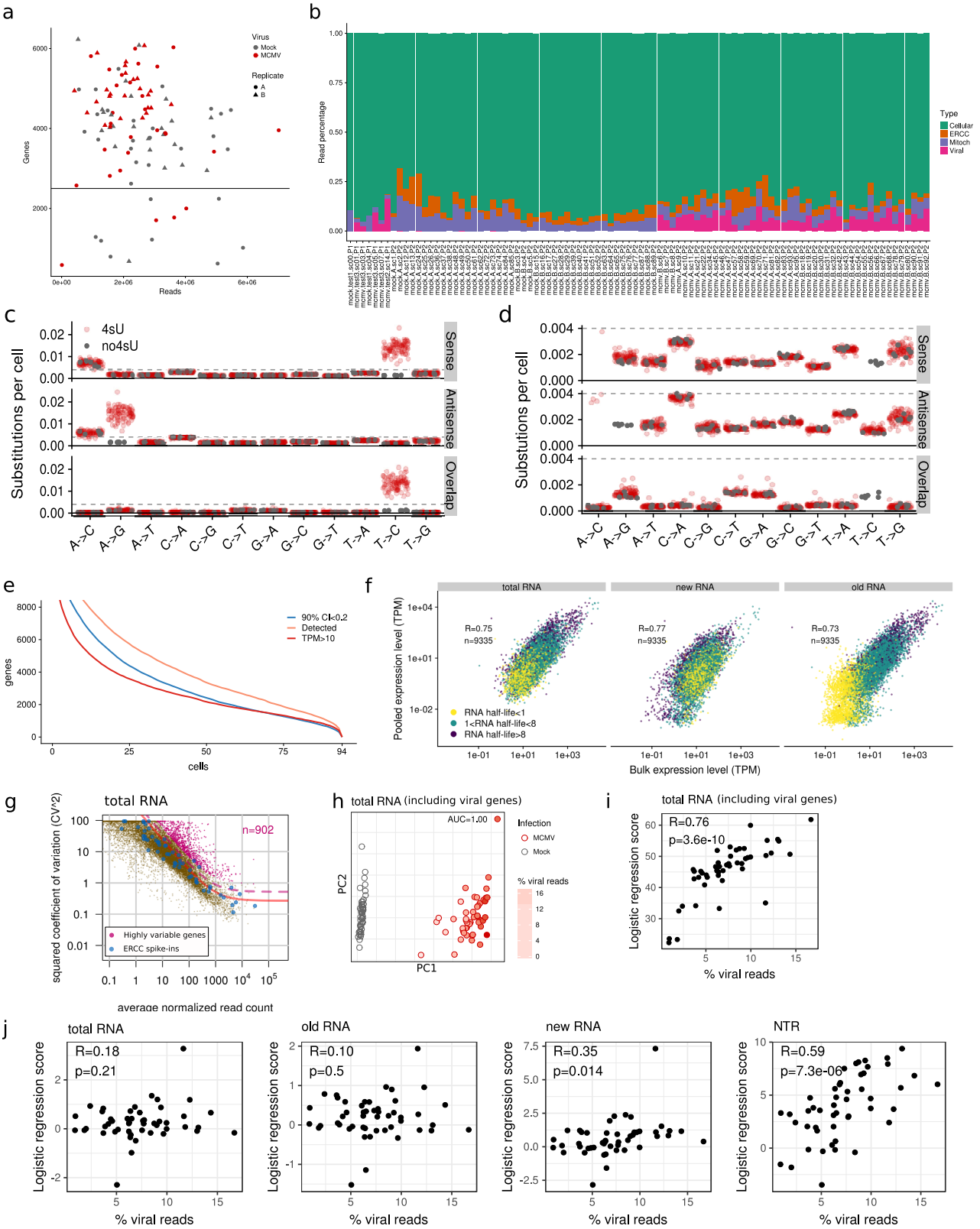
Author contributions Conceptualization: F.E., A.-E.S. and L.D.; computational methodology: M.L., C.S.J. and F.E.; investigation: F.E., M.A.P.B., T.K., T.H., M.L., P.A., F.J.T. and A.-E.S.; infection experiments: M.A.P.B. and T.H.; establishment of scSLAM-seq: M.A.P.B., T.K., P.A., A.-E.S. and L.D; writing: F.E., A.-E.S. and L.D.; funding acquisition: A.-E.S., F.E. and L.D.; supervision: F.J.T., A.-E.S., F.E. and L.D.

Competing interests A patent (EP 18 17 9371) has been filed on the GRAND-SLAM approach to analyse the relative contribution of transcriptional activity based on U-to-C conversions. The authors declare no other competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1369-y.

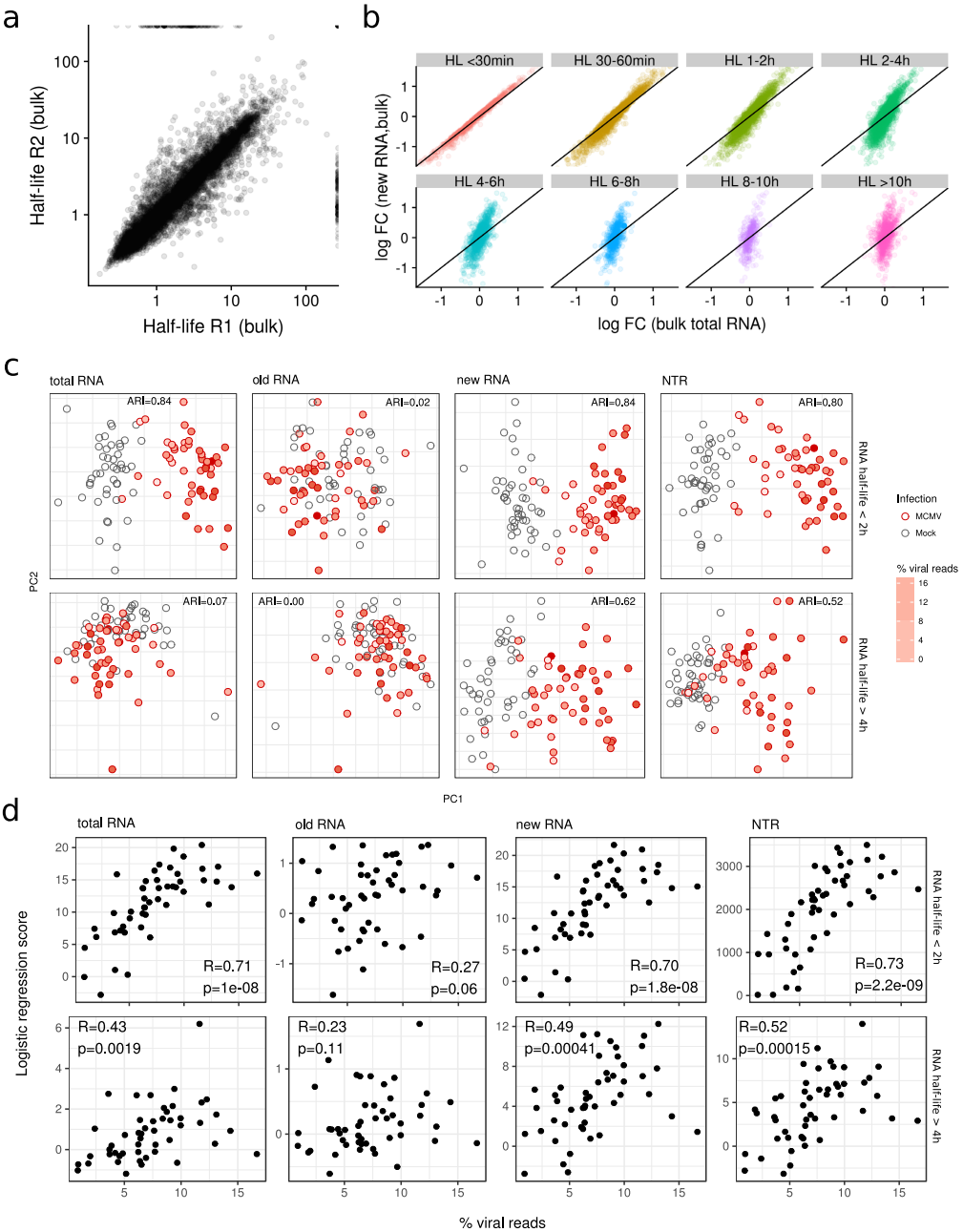
Correspondence and requests for materials should be addressed to F.E. or A.-E.S. or L.D.



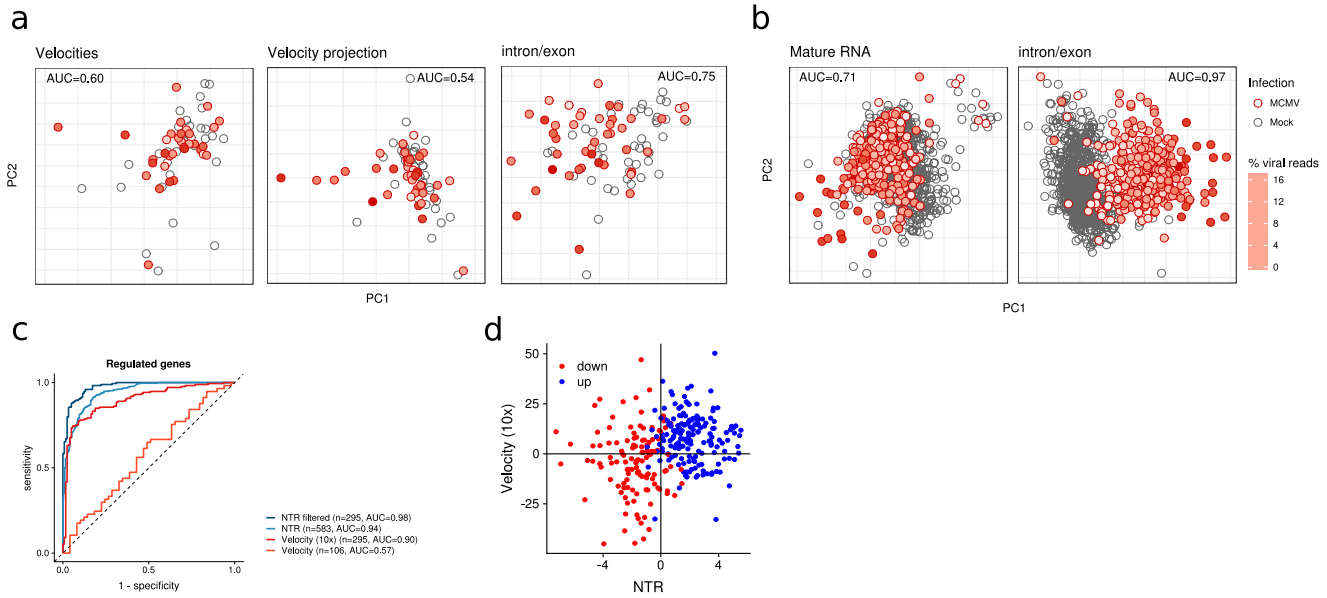
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | scSLAM-seq quality controls. a, Total number of genes detected after scSLAM-seq across all four experimental conditions (uninfected and CMV-infected cells; two biological replicates) versus the total read counts per single cell. The horizontal line indicates a threshold below which cells were excluded from the analysis. **b**, Partition of reads devoted for host (cellular), viral, spike-in control (External RNA Controls Consortium (ERCC)) and mitochondrial genes (Mitoch) across all individual cells. **c**, Rates of nucleotide substitutions demonstrate efficient conversion rates in 4sU-treated single cells (4sU) compared with 4sU-naive cells (no4sU). This was true for reads originating from both cDNA strands (sense and antisense) as well as overlapping parts of the paired-end sequencing (overlap). **d**, As in **c**, zoomed into the range (*y* axis) 0 to 0.004. **e**, Number of genes per cell for which the NTR could be quantified with high precision (90% credible interval (CI) < 0.2) compared with the detected genes and reliably detected genes (TPM >10). **f**, Correlation between expression levels of bulk RNA-seq with the pooled scRNA-seq data for total, new and old RNA. Genes are coloured according to RNA half-life. Pearson's correlation coefficient (R) and the number of genes used (n) are indicated. **g**, Identification of highly variable genes (magenta) using ERCC spike-ins to model the technical noise applied to total RNA (1% false discovery rate). Squared coefficients of variation (CV^2) are plotted against the average normalized read counts for all cells that pass the quality-control filters. The solid pink line fits the average values for ERCC spike-ins (blue dots)[25]. The dashed line marks the expected position of genes with 50% biological coefficient of variation. **h**, PCA (the two first components are depicted) of highly variably genes including the viral transcripts (infected: $n = 2$ replicates, 49 cells; uninfected: $n = 2$ replicates, 45 cells). **i**, Correlation of the percentage of viral reads with the distance to uninfected cells in the first two principal components as measured by logistic regression for the PCA in **f** ($n = 2$ replicates, 49 cells). **j**, As in **i**, for the PCA in Fig. 2a ($n = 2$ replicates, 49 cells).

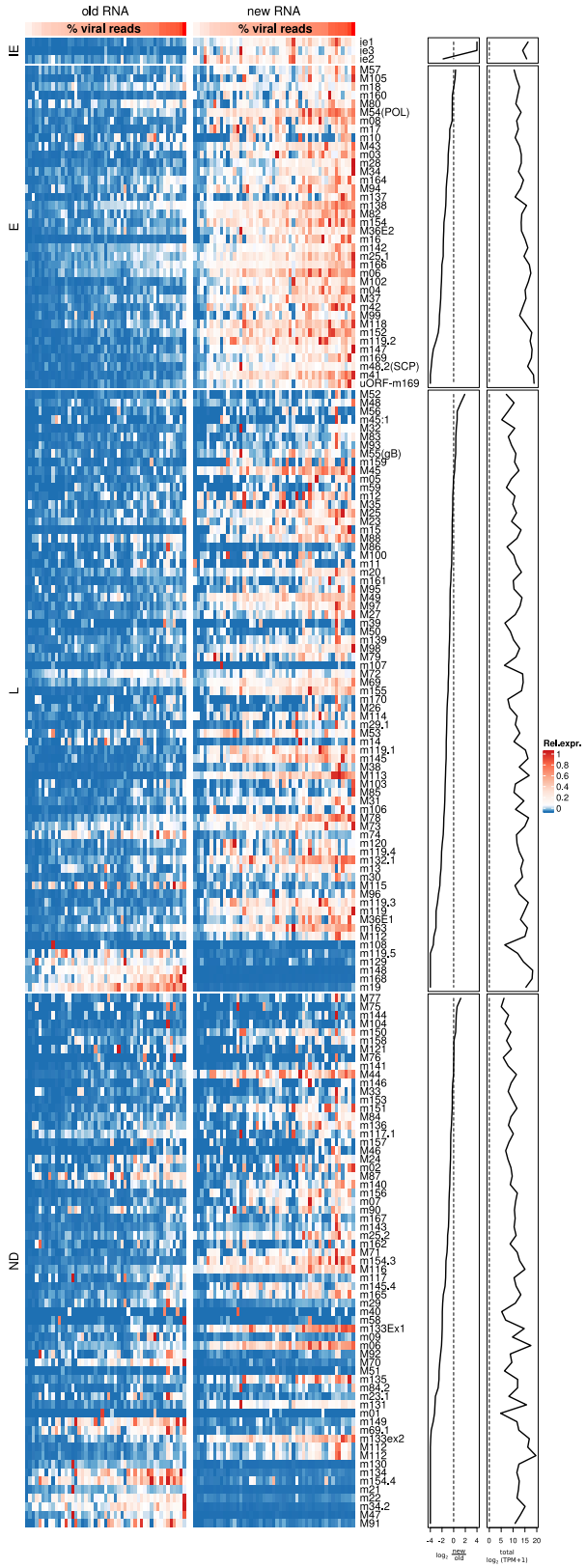
25. Brennecke, P. et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).



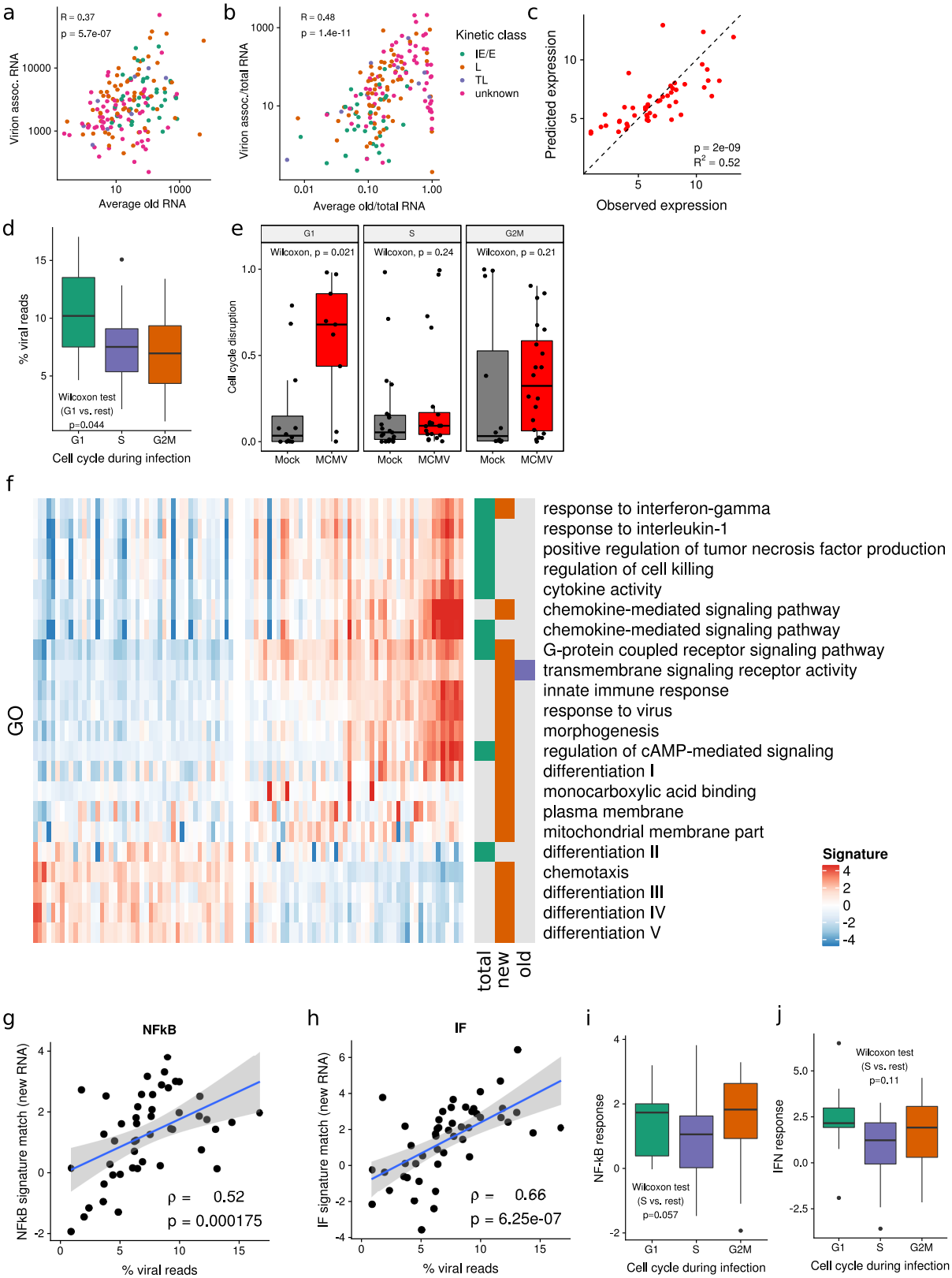
Extended Data Fig. 2 | Half-life estimates and PCA on regulated genes. **a**, Correlation between RNA half-lives estimated from bulk SLAM-seq ($n = 2$ replicates). **b**, The fold change (FC), $\log_2$(MCMV/Mock), of total RNA from bulk SLAM-seq is scattered against the $\log_2$-transformed fold change of new RNA from bulk sequencing, stratified for different RNA half-lives (average of $n = 2$ replicates). **c**, PCA on genes that are differentially expressed in new RNA from the bulk experiments (absolute $\log_2$-transformed fold change > 0.5). Top, PCA on genes with short RNA half-lives (less than 2 h) are shown. Bottom, PCA on genes with long RNA half-lives (more than 4 h). Left to right, PCA was performed using total, old or new RNA, or the NTR (infected, $n = 2$ replicates, 49 cells; uninfected, $n = 2$ replicates, 45 cells). ARI, adjusted rand index. **d**, Correlation analysis of the PCA from **c** with the percentage of viral reads. Pearson's correlation coefficients and *P* values determined by a *t*-test on Pearson's correlation coefficient are indicated (see Extended Data Fig. 1j) ($n = 2$ replicates, 49 cells).



Extended Data Fig. 3 | RNA and NTR velocities. a, PCA computed on velocity values, on expression values projected 1 h into the future using velocity and the intron/exon count ratio for the scSLAM-seq data using the same set of genes as in Fig. 2a (uninfected, $n = 2$ replicates, 43 cells; infected, $n = 2$ replicates, 44 cells). **b**, PCA for the 10x data on the same set of genes as used for Fig. 2a on the basis of mature transcripts (exonic reads only; left) marginally separated uninfected from infected cells. By contrast, on the basis of the ratio of intronic/exonic reads (right), infected cells were almost perfectly separated from uninfected (right; uninfected, $n = 2$ replicates, 793 cells; infected, $n = 2$ replicates, 353 cells). **c**, As in Fig. 2d, except velocities were computed on the basis of degradation rates estimated from uninfected cells only; that is, not violating the steady-state assumption but supplying the class labels (violating the blind test of prediction). **d**, Scatterplot comparing NTRs with velocities from the 10x scRNA-seq data for down- and upregulated genes.

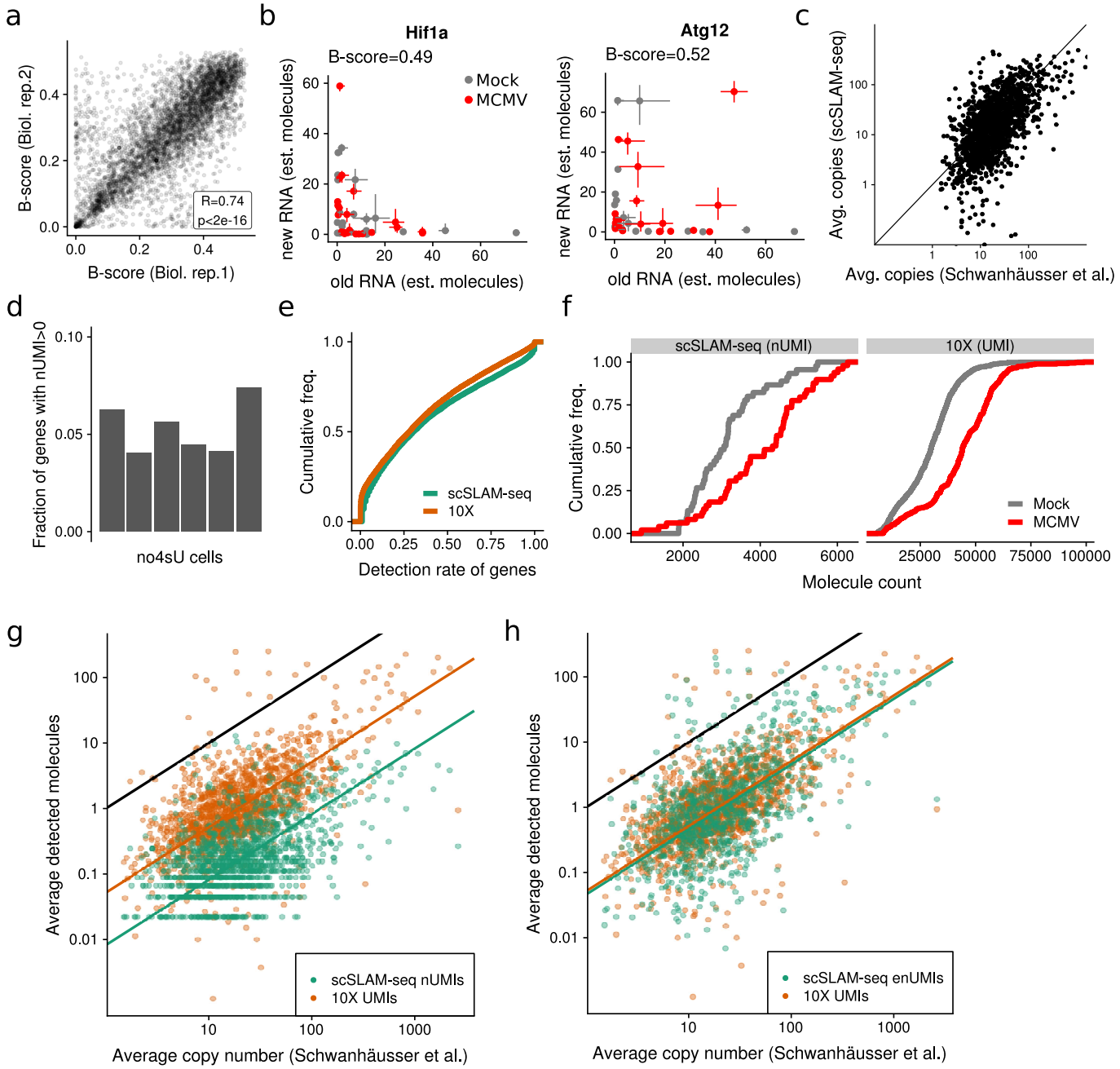


Extended Data Fig. 4 | scSLAM-seq differentiates incoming virion-associated RNA from de novo transcribed viral RNA. Heat maps showing the levels of old (left) and new (right) RNA relative to the maximal total level for each viral gene (rows) per CMV-infected cell (columns). Cells are sorted according to the percentage of viral reads among all reads from the cell. Kinetic classes are indicated. E, early; IE, immediate early; L, late; ND, not defined. The ratio of new to old RNA ($\log_2$(new/old)) and total expression from pooled cells for each viral gene are depicted.



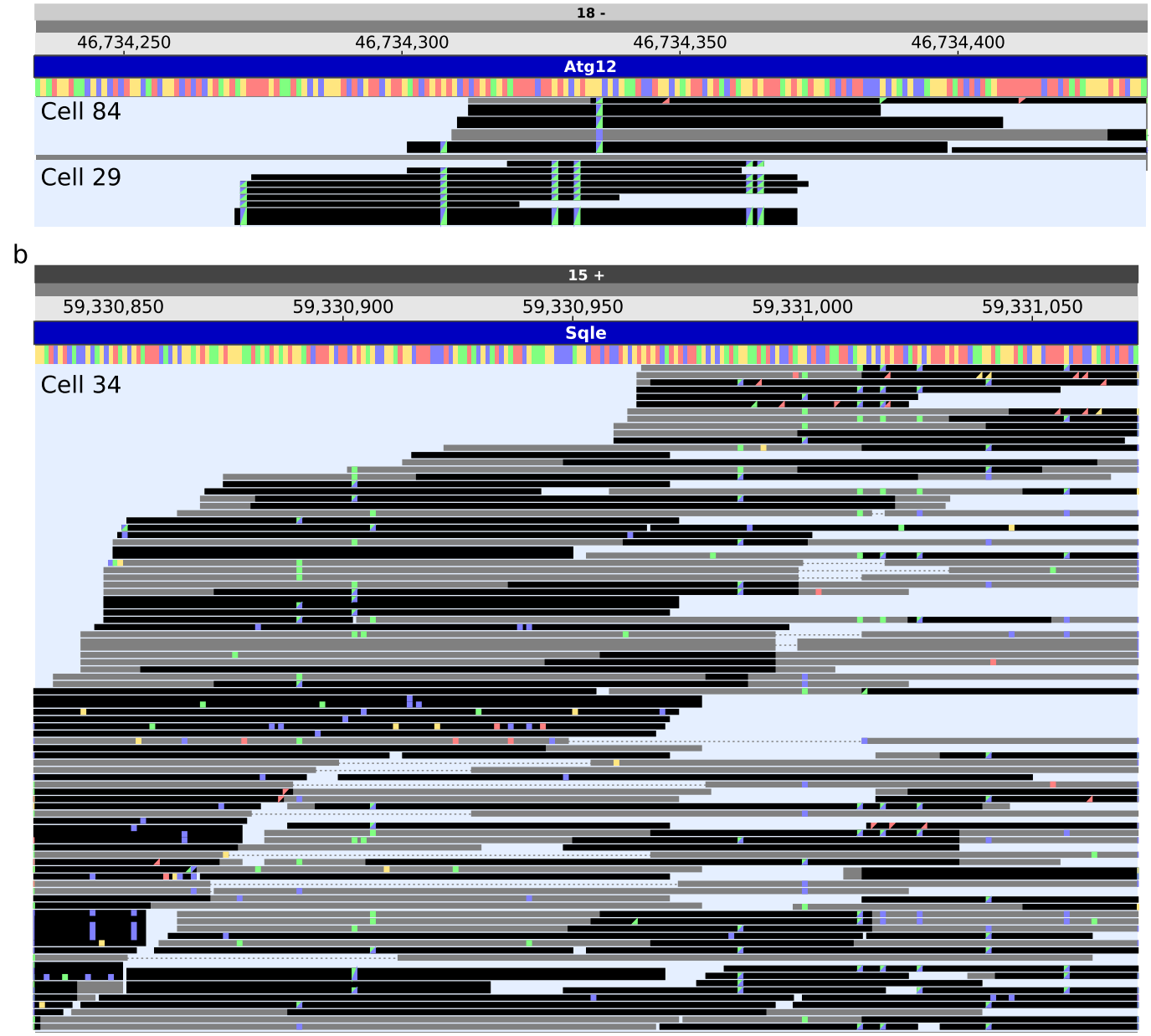
Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Correlation of viral gene expression with dose of infection and cell cycle. a, Comparison of virus stock-derived RNA (virion-associated RNA) and old RNA levels of CMV genes. Mean levels obtained from four independent virus stock vials and mean expression levels in the CMV-infected cells ($n = 2$ replicates, 49 cells) are compared. The colours indicate viral genes of different kinetic classes. TL, true late. Pearson's correlation coefficient and P values determined by a t-test on Pearson's correlation coefficient are indicated. **b**, As in **a** but normalized for the total expression levels of the respective genes in the CMV-infected cells. Pearson's correlation coefficient and P values are indicated. **c**, Scatter plot comparing the predicted extent of viral gene expression per individual cell ($n = 2$ replicates, 49 cells) on the basis of the dose of infection with the observed expression. P value determined by likelihood ratio test. **d**, Distribution of viral reads for cells in G1 ($n = 9$ cells), S ($n = 20$ cells) or G2/M ($n = 20$ cells) phases at the beginning of infection. P value determined by two-sided Wilcoxon test. **e**, Extent of cell-cycle disruption on the basis of cell-cycle projections derived from old and total RNA of uninfected (mock, $n = 2$ replicates, 45 cells) and CMV-infected ($n = 2$ replicates, 49 cells) cells for G1, S and G2/M phases. Individual cells are shown as dots. P values determined by two-sided Wilcoxon tests. **f**, Unbiased pathway and gene set overdispersion analysis (PAGODA)[12] revealed Gene Ontology (GO) terms associated with mock- and CMV-infected cells. The fraction (total, new or old) in which each signature was found is indicated. **g**, **h**, The NF-κB (**g**) and IFN (**h**) response signature score for each cell ($n = 2$ replicates, 49 cells) is plotted against its viral RNA content. The linear regression fit (line), 95% credible interval (shading), Spearman's ρ values and P values determined by t-test on Spearman's ρ are indicated. **i**, **j**, Distribution of the extent of the NF-κB (**i**) and IFN (**g**) responses for cells in G1 ($n = 9$ cells), S ($n = 20$ cells) or G2/M ($n = 20$ cells) phase at the beginning of infection. P values determined by two-sided Wilcoxon tests. All box plots are as in Fig. 4c. Dots represent outliers.

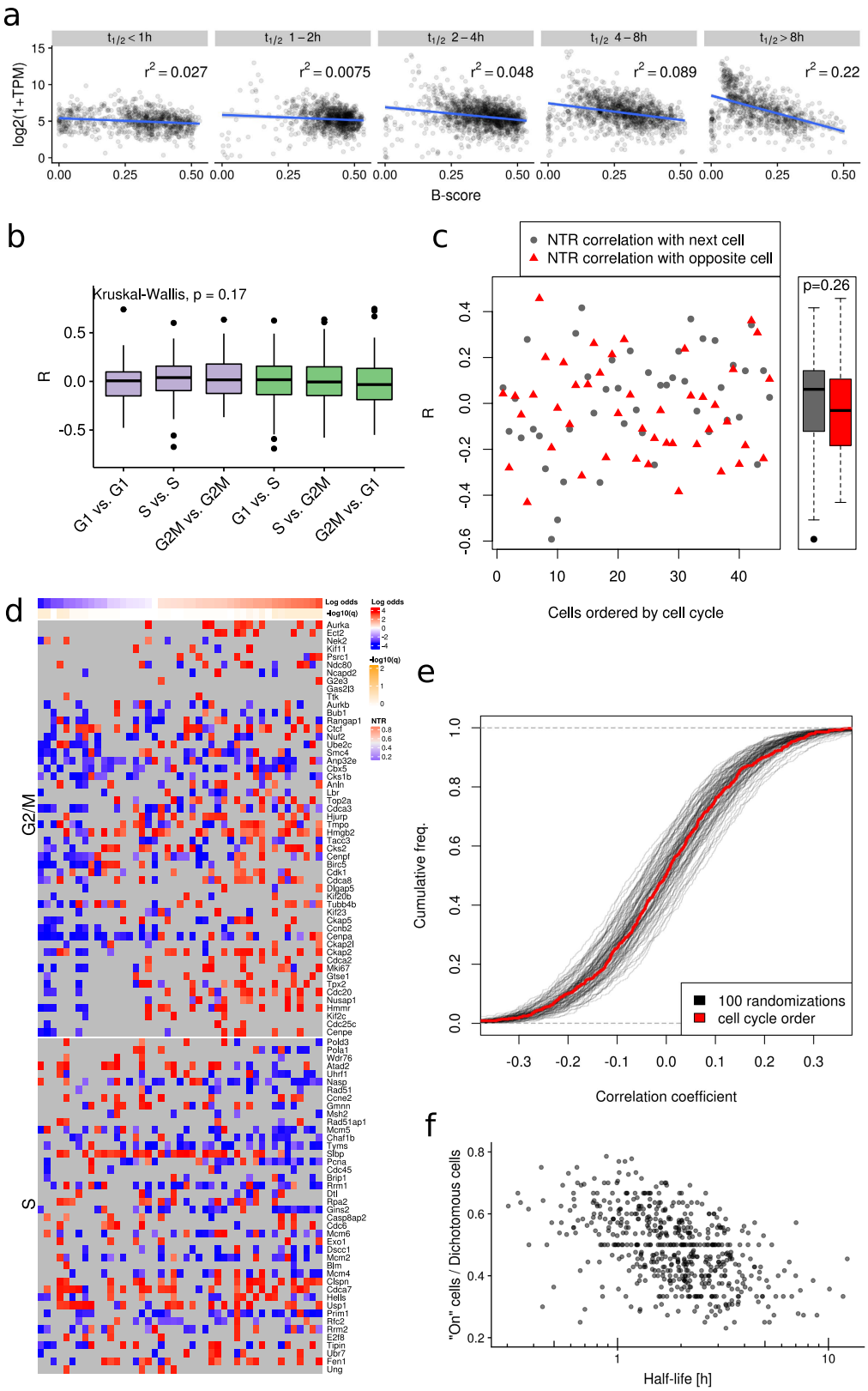


Extended Data Fig. 6 | *B*-scores, nUMIs and transcriptional bursts. **a**, Correlation of *B*-scores ($n = 2$ replicates). Pearson's correlation coefficient and *P* values determined by *t*-test on Pearson's correlation coefficient are indicated. **b**, The number of old and new molecules (estimated by regression analysis with RNA spike-ins) is shown for *Hif1a* and *Atg12*. Both show extreme NTR variance, which results from very few sampled mRNA molecules. Dots represent maximum a posteriori estimates of uninfected (mock, $n = 2$ replicates, 45 cells) and CMV-infected ($n = 2$ replicates, 49 cells) cells. Error bars denote 90% credible intervals provided by GRAND-SLAM. **c**, The average mRNA copy numbers obtained from ref. [26] are scattered against the average copy numbers estimated by regression analysis with RNA spike-ins ($P < 2 \times 10^{-16}$, two-sided *t*-test on Pearson's correlation coefficient; see Supplementary Methods). **d**, The fraction of genes with nUMIs > 0 for all cells with detectable reads is shown for the six samples that were not labelled with 4sU. **e**, The distribution of gene-wise detection rates is shown for all genes that were detected at least once in the 10x and scSLAM-seq data ($n = 12{,}784$). A gene is called detected with at least one UMI or read in the 10x or scSLAM-seq data, respectively. **f**, The distributions of nUMI (left) or UMI (right) counts per cell are shown for uninfected and infected cells in the scSLAM-seq or 10x experiments, respectively. **g**, The average copy number obtained from ref. [26] is plotted against the average number of UMIs and nUMIs from the 10x uninfected ($n = 2$ replicates, 353 cells) and scSLAM-seq ($n = 2$ replicates, 45 cells) experiments, respectively. Lines represent the median capture rates. **h**, As in **f**, but using enUMIs.

26. Schwanhäusser, B. et al. Global quantification of mammalian gene expression control. *Nature* **473**, 337–342 (2011).



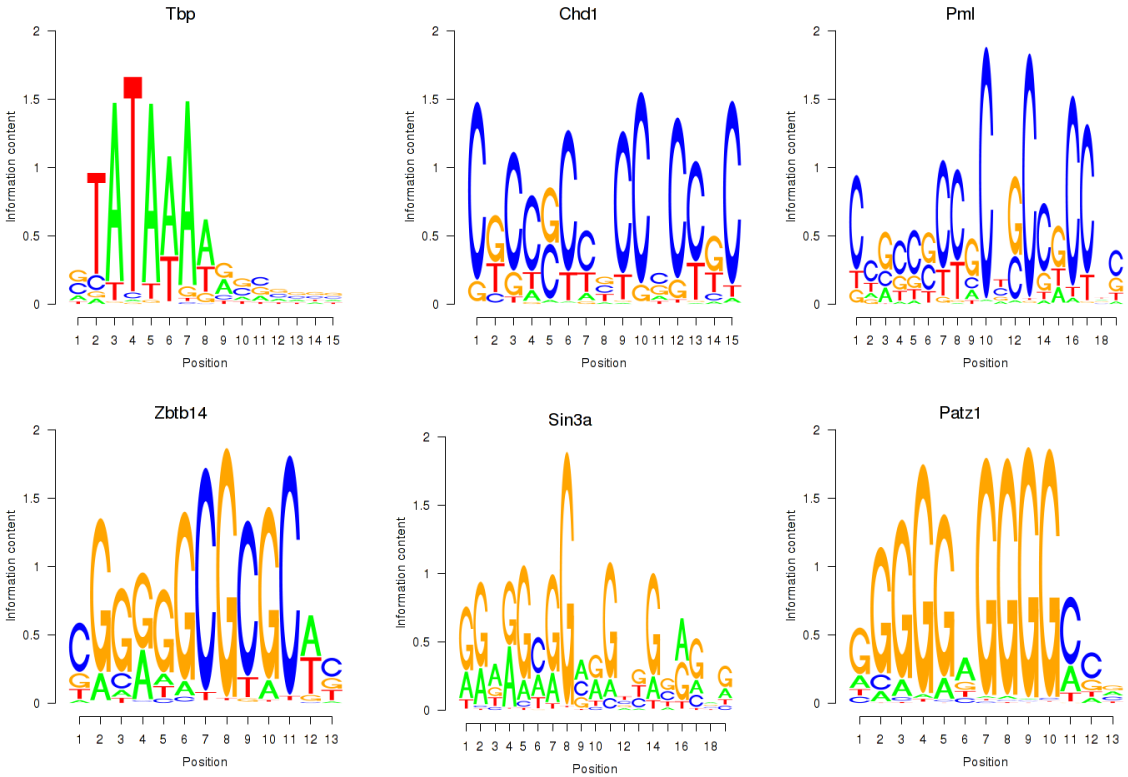
Extended Data Fig. 7 | Genome browser screenshots visualizing nUMIs for exemplary genes. a, Example genome viewer screenshot of two single cells for the *Atg12* gene showing individual reads. Grey and black denote the singly and doubly sequenced parts. The genomic sequence is colour-coded (A, red; C, green; G, blue; T, orange). Mismatches are indicated on the reads with the same colour code. On doubly sequenced parts, the top and bottom triangle represents the corresponding mismatches of the first and second read, respectively. In cell 84, there are two characteristic 4sU mismatches, which are observed in all reads. In cell 29, there are six mismatches, which are distinct from those observed in cell 84. In both cells, only a single new transcript with stochastic 4sU incorporation thus gave rise to the respective reads. **b**, Example genome browser screenshot for the *Sqle* gene. Here, at least four (nUMI = 4) mRNAs gave rise to the observed reads. However, as not all reads overlap, this is likely to be an underestimation of the actual number of cloned transcripts.



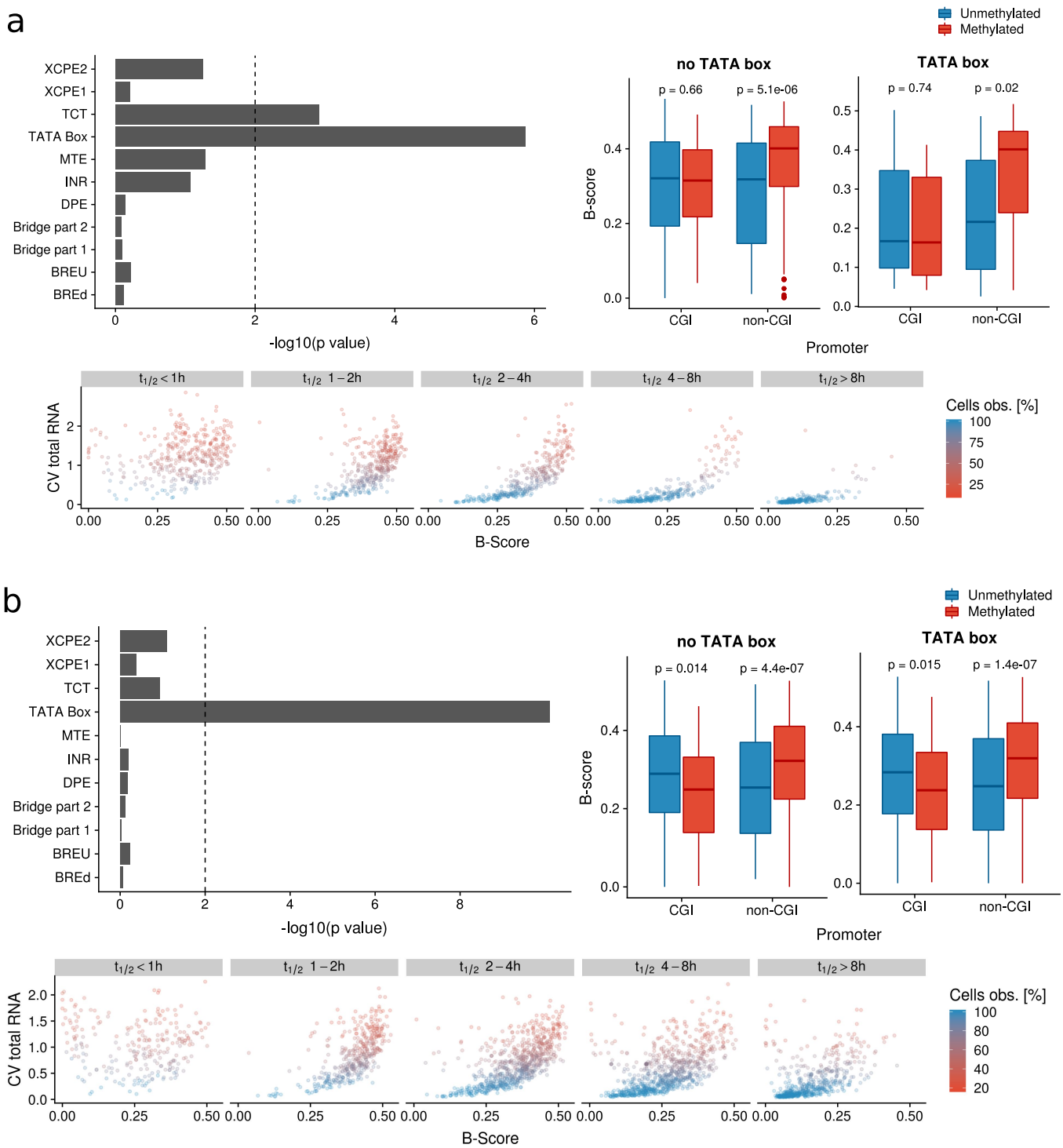
Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | *B*-scores reflect stochastic transcriptional activity in single cells. a, Comparison of *B*-scores (from $n = 2$ replicates, 45 cells) with bulk RNA expression levels stratified by RNA half-life. The r^2 values of ordinary linear regression are shown (lines indicated). Especially for genes with short-lived transcripts, there was no correlation indicating that high *B*-scores are not due to inefficient RNA capture. **b**, Pearson's correlation coefficient of NTR values for the top 10% most-variable genes for pairs of cells ($n = 2$ replicates, 45 cells) either in the same cell-cycle phase (purple), or in different stages of the cell cycle (green). **c**, Cells ($n = 2$ replicates, 45 cells) were ordered according to the cell cycle using reCAT (recover cycle along time)[27] (x axis). The correlation of each cell with the next cell in the order, or the cell farthest away in the order (opposite), is shown. Pearson's correlation coefficients were computed on the NTR values of the top 10% most-variable genes. **d**, Heat map showing the NTR values for marker genes of the S and G2/M phases of the cell cycle. Grey fields indicate undetected genes. Cells (columns, $n = 2$ replicates, 45 cells) were ordered according to the log odds (on–off versus S–G2/M). The log odds and associated *P* values (two-sided Fisher's exact test, corrected by Benjamini–Hochberg) are indicated. Genes (rows) were ordered according to the correlation of their NTR values with the log odds order. **e**, Distribution of the Spearman correlation coefficient of the top 10% most-variable genes (*B*-score against cell-cycle log-odds order; see **d**). For the sake of comparison, cells were permuted randomly 100 times, and the corresponding distributions of the correlation coefficient are indicated. **f**, Scatter plot of RNA half-lives and the fraction of on cells ($n = 2$ replicates, 45 cells) among all cells with the gene in either on or off state. All box plots are as in Fig. 4c. Dots denote outliers.

27. Liu, Z. et al. Reconstructing cell cycle pseudo time-series via single-cell transcriptome data. *Nat. Comm.* **8**, 22 (2011).



Extended Data Fig. 9 | Sequence logos overrepresented in promoter regions of dichotomous genes. Sequence logos of the transcription factors with significantly enriched binding sites among genes with low (*Tbp*) or high (*Patz1*, *Pml*, *Chd1*, *Sin3a* and *Zbtb14*) *B*-scores obtained from the SwissRegulon database[13].



Extended Data Fig. 10 | Bursting kinetics analyses repeated for subsets of genes. a, Analyses as in Fig. 4a–e repeated only for the 1,718 significantly regulated genes according to the heterogeneity test (on the basis of nUMIs and enUMIs) (see Supplementary Methods). **b**, Analyses as in Fig. 4a–e repeated for genes with the top 50% ($n = 2{,}770$) of expressed genes. All box plots are as in Fig. 4c. *P* values were determined by two-sided Wilcoxon test.

Corresponding author(s): Lars Dölken, Antoine-Emmanuel Saliba, Florian Erhard

Last updated by author(s): 01-02-2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Next-generation sequencing data was collected by Illumina NextSeq500 (bulk RNA-seq, scSLAM-seq) or NovaSeq 6000 (10x Chromium libraries). Data was demultiplexed using bcl2fastq2 Conversion Software v.2.
Data analysis	Trimmomatic (0.36), STAR (2.5.3a), Cell Ranger (2.1.1), GRAND-SLAM (2.0), velocyto (0.17.7), Scanpy (1.0.4), GOrilla, R (3.4.4) with packages: SC2P (1.0.2), scde (1.99.1), Velocyto.R (1.0.6), reCAT (github commit 1d091de), scran (1.6.9), lfc (0.2.1). Script files to generate all figures are available at zenodo (doi: 10.5281/zenodo.1299120).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The sequencing data are available at Gene Expression Omnibus (GEO) at GSE115612.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We performed scSLAM-seq on 107 single murine fibroblast cells and in parallel analyzed global transcriptional changes of a matched larger (1x10^5) population of cells using SLAM-seq. We aimed to sequence about 50 cells per condition to get reproducible gene expression data (from at least 5 cells that express the gene) and analyze transcriptional activity of genes which were transcribed in only 10% of cells. We performed a 10x Chromium droplet-based scRNA-seq run to compare droplet-based scRNA-seq data with our scSLAM-seq approach. The respective run provided us with hundreds of uninfected (n=793) and MCMV-infected (n=353). This was clearly sufficient to compare to the 49 CMV-infected and 45 uninfected cells.
Data exclusions	For scSLAM-seq, we filtered out cells with less 2,500 quantified genes (TPM>1) (Extended Data Fig. 1a). Using a cut-off of 2,500 quantified genes per single cell is a well-established cut-off in the field of scRNA-seq and simply ensures that only cells with decent coverage rates of their transcriptome are included into the subsequent analyses.
Replication	During the establishment of scSLAM-seq, we first prepared scSLAM-seq libraries from 50, 5 and 1 cell. Only once this was succesful, we analzyed two biological replicates (Extended Data Fig. 1a). These reproduced and confirmed the findings from both the smaller number of cells, cell pools and bulk sequencing data.
Randomization	No randomization was required for this study as samples only comprised uninfected and infected cells.
Blinding	Investigators were not blinded during data collection and analysis as this would not have been possible during sample preparation, data collection and analysis. Furthermore, blinding was not relevant to this study. Only scSLAM-seq and the analysis of new RNA enables us to differentiate the infected from the uninfected cells.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Murine NIH-3T3 fibroblasts were obtained from ATCC (ATCC® CRL-1658).
Authentication	Not authenticated but can be deduced from RNA-seq data.
Mycoplasma contamination	Cell culture was negatively tested for Mycoplasma contamination by PCR.
Commonly misidentified lines (See ICLAC register)	None of the cell lines used in this study is listed in the ICLAC register.

LETTER

https://doi.org/10.1038/s41586-019-1374-1

m^{6}A enhances the phase separation potential of mRNA

Ryan J. Ries[1], Sara Zaccara[1], Pierre Klein[1], Anthony Olarerin-George[1], Sim Namkoong[2], Brian F. Pickering[1], Deepak P. Patil[1], Hojoong Kwak[3], Jun Hee Lee[2] & Samie R. Jaffrey[1]*

N^6-methyladenosine (m^{6}A) is the most prevalent modified nucleotide in mRNA[1,2], with around 25% of mRNAs containing at least one m^{6}A. Methylation of mRNA to form m^{6}A is required for diverse cellular and physiological processes[3]. Although the presence of m^{6}A in an mRNA can affect its fate in different ways, it is unclear how m^{6}A directs this process and why the effects of m^{6}A can vary in different cellular contexts. Here we show that the cytosolic m^{6}A-binding proteins—YTHDF1, YTHDF2 and YTHDF3—undergo liquid–liquid phase separation in vitro and in cells. This phase separation is markedly enhanced by mRNAs that contain multiple, but not single, m^{6}A residues. Polymethylated mRNAs act as a multivalent scaffold for the binding of YTHDF proteins, juxtaposing their low-complexity domains and thereby leading to phase separation. The resulting mRNA–YTHDF complexes then partition into different endogenous phase-separated compartments, such as P-bodies, stress granules or neuronal RNA granules. m^{6}A-mRNA is subject to compartment-specific regulation, including a reduction in the stability and translation of mRNA. These studies reveal that the number and distribution of m^{6}A sites in cellular mRNAs can regulate and influence the composition of the phase-separated transcriptome, and suggest that the cellular properties of m^{6}A-modified mRNAs are governed by liquid–liquid phase separation principles.

To understand how m^{6}A affects the fate of mRNA, we considered the biochemical properties of the major cytosolic m^{6}A-binding proteins YTHDF1, YTHDF2 and YTHDF3 (hereafter denoted DF1, DF2 and DF3, respectively). These paralogous proteins have high sequence identity and comprise a YTH domain of around 15 kDa that binds m^{6}A, and a low-complexity domain of around 40 kDa that includes prion-like domains[4] (Extended Data Fig. 1a).

Some low-complexity amino acid sequences form fibrils, hydrogels or liquid droplets as a result of phase separation[5,6]. To test whether DF

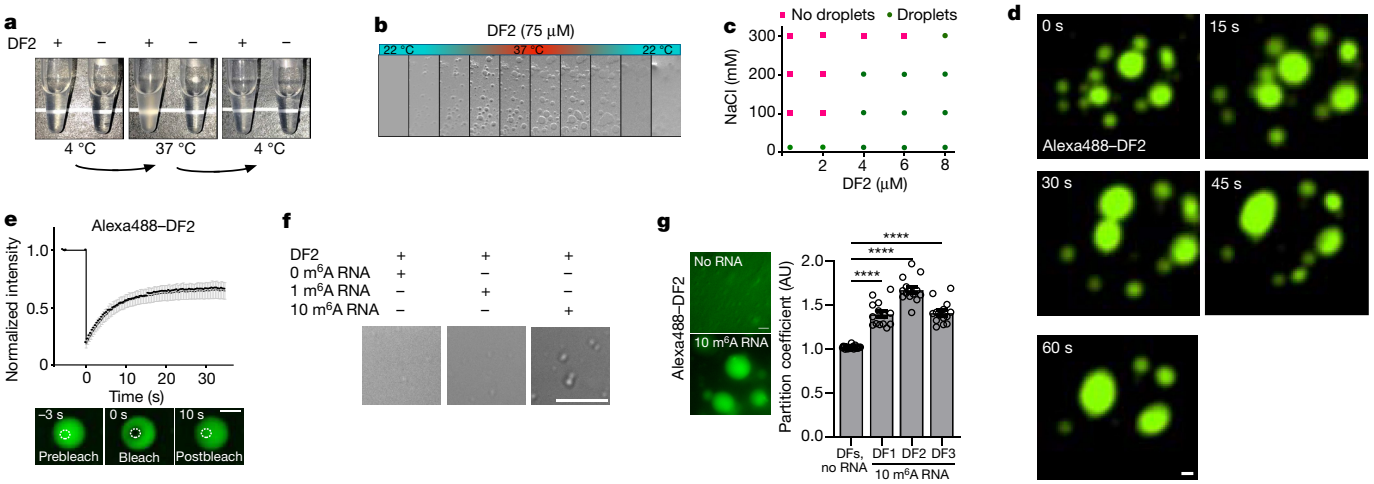


Fig. 1 | Polymethylated m^{6}A RNAs trigger liquid–liquid phase separation of DF proteins. a, Tubes containing either buffer only or recombinant DF2 were heated from 4 °C to 37 °C. DF2 undergoes phase separation when heated; this is reversible upon cooling. **b**, Time-lapse of bright-field microscopy images of DF2 droplets (75 μM) subjected to a temperature gradient. The temperature was increased at a rate of 1 °C per minute from 22 °C to 37 °C, enabling the formation of protein droplets. Lowering the temperature back to 22 °C causes disassembly of the droplets. **c**, Phase diagram of DF2 in the presence of different concentrations of NaCl, showing that salt dampens the phase-separation potential of the protein. Green circles indicate that protein droplets were present; pink squares indicate that no protein droplets were observed in the buffer. **d**, Alexa488–DF2 (75 μM) was imaged by fluorescence microscopy over 1 min. The fusion of Alexa488–DF2 droplets can be seen in Supplementary Video 1. **e**, Top, changes in the fluorescence intensity of Alexa488–DF2 droplets after photobleaching were plotted over time. The background was subtracted from the fluorescence measurement. The black curve represents the mean of the fluorescence intensity in the photobleached region of interest in distinct droplets ($n = 8$); grey bars indicate s.e.m. Bottom, representative images of fluorescence recovery. **f**, A 65-nucleotide RNA containing 10 m^{6}A nucleotides (570 nM) induces DF2 (25 μM) to rapidly form small liquid droplets, whereas RNA containing one m^{6}A nucleotide or no m^{6}A nucleotides does not cause substantial DF2 phase separation. **g**, Left, the addition of RNA containing 10 m^{6}A sites (425 nM) enhances the phase separation of DFs (15 μM) in solution. Right, partition coefficients (PCs) for DFs in the presence and absence of RNA containing 10 m^{6}A sites. For the no-RNA condition, partition coefficients were calculated immediately before the addition of m^{6}A-containing RNA (right panel: DFs with no RNA, mean PC = 1.0; DF1, $n = 8$; DF2, $n = 10$; DF3, $n = 9$; total, $n = 27$). Partition coefficients for the DF proteins were measured shortly after the addition of RNA containing 10 m^{6}A nucleotides and the mean DF partition coefficients increased notably (right panel: DF1 mean PC = 1.40, $n = 14$; DF2 mean PC = 1.67, $n = 14$; DF3 mean PC = 1.41, $n = 14$ droplets) within minutes of RNA addition. Error bars represent s.e.m. n represents the number of droplets from technical replicates. ****$P < 0.0001$, two-sided Mann–Whitney test. Scale bars, 10 μm (**d**–**g**).

[1]Department of Pharmacology, Weill-Cornell Medical College, Cornell University, New York, NY, USA. [2]Department of Molecular and Integrative Physiology, University of Michigan, Ann Arbor, MI, USA. [3]Department of Molecular Biology and Genetics, Cornell University, Ithaca, NY, USA. *e-mail: srj2003@med.cornell.edu

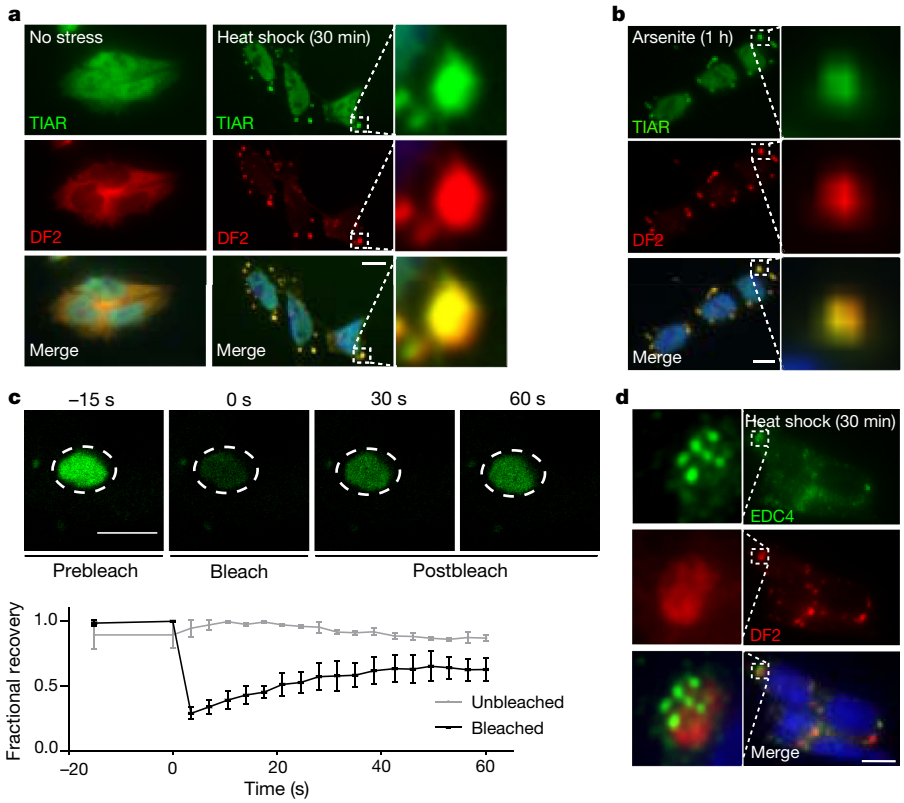


Fig. 2 | DF proteins exhibit liquid-like properties in cells and relocalize during stress. a, **b**, Co-immunostaining of DF2 (red) and the stress-granule marker TIAR (green) in mES cells before and after incubation at 42 °C for 30 min (**a**) or treatment with sodium arsenite (0.5 mM) for 1 h (**b**). DF2 relocalizes to stress granules, as visualized by its co-localization with TIAR (yellow) in the overlay panel (bottom). **c**, NeonGreen–DF2 was endogenously expressed in HEK293 cells using CRISPR–Cas9 knock-in and treated with arsenite (0.5 mM, 1 h). NeonGreen–DF2 partitioned into arsenite-induced stress granules. Photobleaching of stress granules is followed by rapid recovery of fluorescence, indicating that NeonGreen–DF2 can actively undergo phase separation in cells. The line traces represent mean fractional fluorescence (unbleached, $n = 3$; bleached, $n = 3$), error bars represent s.e.m. **d**, P-bodies have been shown to be adjacent to stress granules[29]. We observed the proximity between P-bodies and stress granules by co-immunostaining of the stress granule marker DF2 (red) and the P-body marker EDC4 (green) in mES cells after heat-shock stress (42 °C, 30 min). DF2-labelled stress granules and P-bodies are in close proximity but do not co-localize. Scale bars, 10 μm (a, b, d) and 5 μm (c).

proteins form these condensates, we purified full-length recombinant DF2, which is the most abundant DF paralogue in most cells[4]. Solutions of DF2 were clear at 4 °C, became turbid upon warming to 37 °C, and then became clear again after cooling to 4 °C (Fig. 1a). Using phase-contrast microscopy, we observed protein droplets that formed only in the warmed samples (Fig. 1b). This warming-induced liquid–liquid phase separation (LLPS) is suggestive of a lower critical solution temperature phase separation[7]. This type of phase separation is associated with Pro-X_n-Gly motifs, which are enriched in DF proteins (Extended Data Fig. 1a).

LLPS of DF2 is enhanced by increased protein concentration and dampened in the presence of sodium chloride (Fig. 1c). The addition of as little as 10% glycerol and lowering the concentration of NaCl reduced the concentration of DF2 required for the phase transition to between 1 μM and 8 μM (Fig. 1c). These values are consistent with the intracellular concentration of endogenous DF proteins[8] (around 5 μM). Through imaging studies of Alexa Fluor 488-labelled DF2 (Extended Data Fig. 1b), we observed droplets fusing to form larger droplets (Fig. 1d, Supplementary Video 1). Photobleaching of a region of an Alexa488-labelled DF2 droplet was associated with rapid recovery of fluorescence (Fig. 1e), which is consistent with DF2 exhibiting liquid-like properties[9]. The low-complexity domain of DF2 is required for LLPS, as removal of this domain prevented LLPS from occurring even at high protein concentrations (Extended Data Fig. 1c). Each DF paralogue was found to exhibit LLPS (Extended Data Fig. 1d), and mixing all three DF proteins resulted in droplets that contained all three proteins, which suggests that these proteins interact and undergo phase separation together (Extended Data Fig. 1e). Overall, these studies reveal that LLPS is a physical property of the DF proteins, at least in vitro.

We next asked whether the binding of m^{6}A-RNA to the YTH domain regulates LLPS of DF proteins. We chose a buffer and protein concentration in which LLPS does not occur. Addition of a 65-nucleotide-long RNA containing either zero m^{6}A nucleotides or one m^{6}A nucleotide did not induce LLPS of DF2 (Fig. 1f, Extended Data Fig. 1f). However, an RNA that contained ten m^{6}A nucleotides triggered LLPS within minutes (Fig. 1f, Extended Data Fig. 1g, Supplementary Video 2) and increased the partition coefficient of each DF protein (Fig. 1g, Extended Data Fig. 1h). These data suggest that polymethylated m^{6}A-RNA provides a scaffold that juxtaposes several DF proteins, causing them to undergo LLPS through interactions between their low-complexity domains. The liquid droplets are therefore composed of a DF–RNA coacervate.

DF proteins localize to neuronal RNA granules, P-bodies and stress granules[4,10,11], each of which is considered to be a phase-separated compartment in the cytosol[4,10,11]. This raises the possibility that LLPS may govern the localization of DF proteins, and potentially of m^{6}A-mRNA. To determine whether DF proteins undergo phase separation in cells, we first examined stress granules because they can be induced by various stimuli in a temporally controlled manner. Diverse stimuli—including heat shock, sodium arsenite and endoplasmic reticulum stress—caused all three DF proteins to relocalize from throughout the cytosol to stress granules in a range of cell types (Fig. 2a, b, Extended Data Fig. 2a–f).

To address whether DF2 exhibits liquid-like properties in vivo, we used CRISPR–Cas9 to insert NeonGreen into the genomic *YTHDF2* locus of HEK293 cells, resulting in the endogenous expression of NeonGreen-labelled DF2 (NeonGreen–DF2) (Extended Data Fig. 2g). Photobleaching of sodium arsenite-induced stress granules resulted in a rapid recovery of NeonGreen–DF2 fluorescence (Fig. 2c, Extended Data Fig. 2h), which is consistent with the liquid-like behaviour of DF2 in vitro.

In non-stressed cells, DF2 is localized to P-bodies[12]. However, after heat-shock stress, we noticed that P-bodies lacked DF2 and that they were instead often located adjacent to DF2-labelled granules (Fig. 2d). This suggests that DF2 can partition into different structures: into P-bodies in unstressed cells, and into stress granules during stress.

A previous study[13] observed that, two hours after heat shock, there was a marked increase in the amount of DF2, which was accompanied by its relocation to the nucleus. However, in our experiments we detected no obvious change in the amount of DF2, which was located exclusively in cytosolic stress granules (Extended Data Fig. 2i). Nevertheless, to determine whether the LLPS of DF2 is due to increased

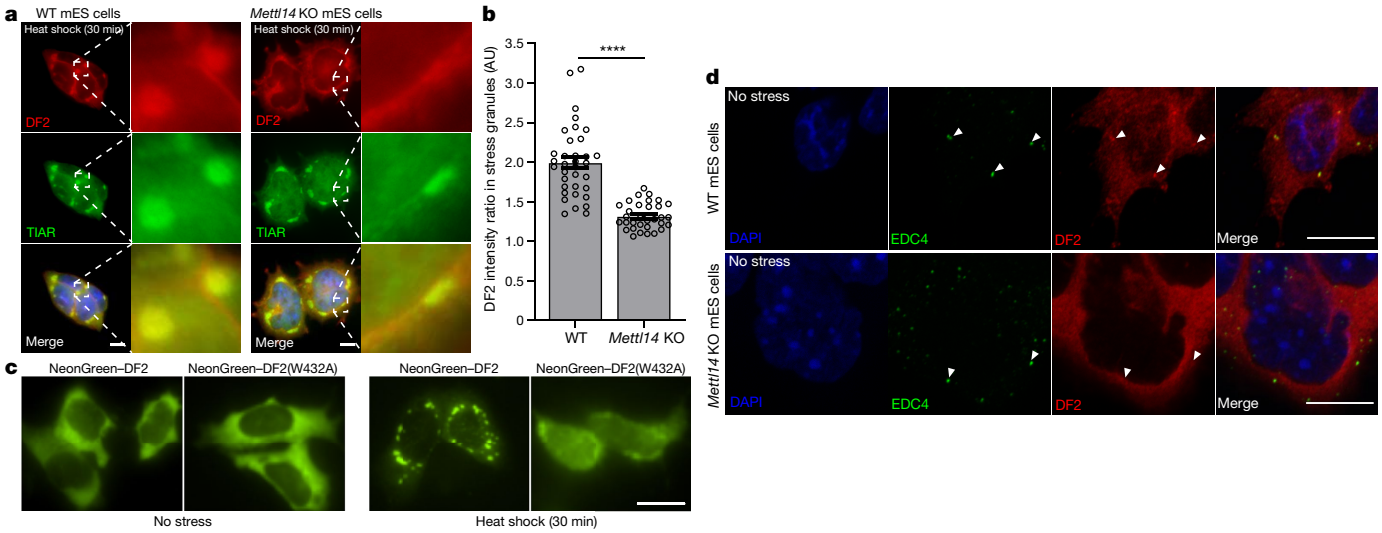


Fig. 3 | m^{6}A enhances the ability of DF proteins to partition into intracellular phase-separated compartments. a, Stress granules form normally in both in wild-type (WT) and *Mettl14*-knockout (KO) mES cells, which lack m^{6}A-mRNA; however, DF2 relocalization in *Mettl14*-knockout mES cells is delayed. Co-immunostaining was performed using the stress-granule marker TIAR (green) and DF2 (red) after heat shock (42 °C, 30 min) or arsenite stress (0.5 mM, 30 min). **b**, DF2 fluorescence intensity ratios in stress granules (DF2 intensity inside TIAR-stained granules/DF2 intensity in the cytoplasm immediately adjacent to TIAR-stained granules) in wild-type and *Mettl14*-knockout mES cells show delayed DF2 co-localization in *Mettl14*-knockout cells. Wild type, $n = 35$; *Mettl14*-knockout, $n = 32$; where n represents stress granules from biological replicates. The height of the bar represents mean fluorescence intensity ratios and the error bars represent s.e.m. $^{****}P < 0.0001$, two-sided Mann–Whitney test. **c**, The localization of a DF2 mutant (DF2(W432A)) with an approximately tenfold reduced affinity for m^{6}A[30] to stress granules is impaired after heat shock (42 °C, 30 min). The W432A mutation disrupts the m^{6}A-binding tryptophan cage in DF2[30]. Plasmids expressing NeonGreen–DF2 and NeonGreen–DF2(W432A) were transfected into wild-type mES cells and images were taken before (left) and after (right) heat shock (42 °C, 30 min). **d**, Co-immunostaining showed well-defined overlap between DF2 (red) and P-bodies as labelled by EDC4 (green) in wild-type mES cells. However, in *Mettl14*-knockout cells, this co-localization was markedly reduced and DF2 appeared more diffusely cytosolic. Representative images from slices of a confocal *Z*-stack are shown. Individual P-bodies and their region of overlap with DF2 are indicated by white arrowheads. Scale bars, 10 μm (a, c, d).

DF2 expression, we performed an experiment using translation inhibitors. The use of these inhibitors did not prevent stress-induced relocalization of DF2 to stress granules (Extended Data Fig. 2j).

We also considered the possibility that stress increases m^{6}A levels in mRNAs. Heat shock and exposure to arsenite can result in increased mRNA methylation when assayed up to 6 h after cell stress[13–16]. Although our assays were performed immediately after stress, we asked whether increased formation of m^{6}A mediates LLPS. Because m^{6}A formation occurs co-transcriptionally[17,18], we used the transcription inhibitor actinomycin D to block m^{6}A formation. However, transcription inhibition did not reduce the localization of DF2 to stress granules (Extended Data Fig. 2j). Additionally, m^{6}A levels in cellular poly(A) mRNA did not change after stress (Extended Data Fig. 2k, l). Overall, no new DF2 protein or new mRNA methylation is needed for DF2 to partition into stress granules. Instead, the existing m^{6}A distribution in mRNA at the onset of stress is sufficient to guide the LLPS of the DF–m^{6}A-mRNA complexes.

Nearly all of the m^{6}A formation in mRNA is catalysed by the METTL3–METTL14 heterodimeric methyltransferase[19–21]. m^{6}A is not required for the formation of stress granules, because stress-granule formation appeared largely normal in *Mettl14*-knockout mouse embryonic stem (mES) cells (Fig. 3a, Extended Data Fig. 3a, b). However, the relocalization of DF2 to stress granules was markedly reduced in *Mettl14*-knockout cells (Fig. 3a, b, Extended Data Fig. 3b). Similarly, a DF2 mutant that does not bind m^{6}A showed reduced relocalization to stress granules in wild-type cells (Fig. 3c). Thus, DF2 must bind m^{6}A-mRNA in order to efficiently partition into stress granules.

We also examined whether m^{6}A-mRNA is required for DF2 localization to P-bodies. P-bodies were readily detected in wild-type and *Mettl14*-knockout mES cells using the P-body marker EDC4 (Fig. 3d). However, in *Mettl14*-knockout cells, DF2 was diffusely cytosolic with no clear P-body enrichment (Fig. 3d). Thus, DF2 is guided to P-bodies by forming complexes with m^{6}A-mRNA.

To further determine whether polymethylated m^{6}A RNAs promote LLPS in cells, we measured m^{6}A levels in mRNA that was purified from stress granules induced by heat shock or by arsenite stress (Fig. 4a, Extended Data Fig. 4a, b). In both cases, m^{6}A levels in stress-granule mRNA were around 45% to 50% higher than in total cellular mRNA.

We also examined the transcriptomes of various RNA granules. We first examined biotin-isoxazole-induced RNA granules from mouse brain extracts, which resemble neuronal RNA granules[22]. In the previous transcriptomic analysis of these structures, the $\log_2$-transformed fold enrichment for each mRNA was reported when it was greater than 1. We classified each mRNA on the basis of the number of mapped m^{6}A sites. mRNAs with no mapped m^{6}A sites showed the lowest enrichment, whereas mRNAs with more mapped m^{6}A sites showed correspondingly higher enrichment (Extended Data Fig. 4c). Thus, polymethylated RNAs exhibit the highest enrichment in these RNA granules.

We observed a similar effect for arsenite-induced stress granules prepared from U2OS cells[23]. mRNAs with zero mapped m^{6}A sites or one mapped m^{6}A site showed no substantial enrichment in stress granules (Fig. 4b). However, for mRNAs with two or more m^{6}A sites, the degree of enrichment in stress granules increased in proportion to the number of mapped m^{6}A sites (Fig. 4b). A similar trend was seen using a transcriptomic analysis of stress granules that were induced with heat shock, thapsigargin, and arsenite in mouse embryonic fibroblast NIH3T3 cells[24] (Extended Data Fig. 4d). Although transcript length is correlated with increased enrichment in stress granules[23,24], the number of m^{6}A sites correlates with stress-granule enrichment even when transcript length is controlled for (Extended Data Fig. 4e). Single-molecule fluorescence in situ hybridization (smFISH) showed that m^{6}A-containing mRNAs exhibit higher levels of stress-granule enrichment than non-methylated mRNAs (Fig. 4c, d). Overall, these data show that polymethylated mRNAs—but not singly methylated RNAs—are enriched in stress granules. Notably, some mRNAs that lack

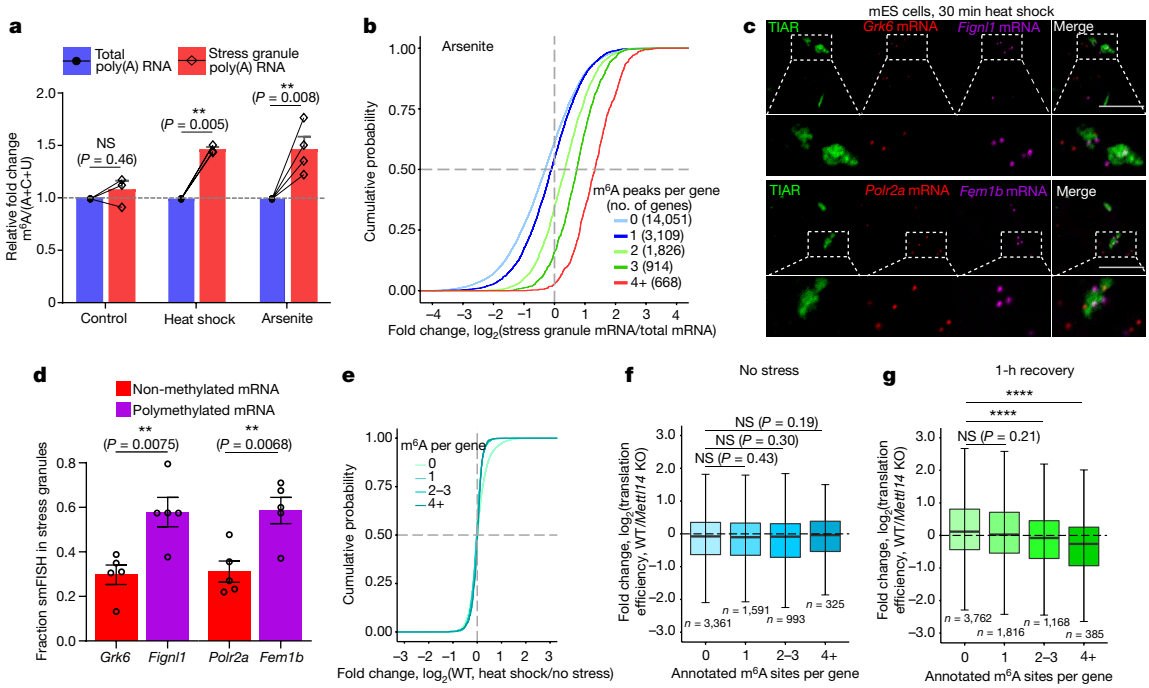


Fig. 4 | m^{6}A-containing mRNAs are enriched in distinct DF-containing RNA granules. a, m^{6}A levels were measured in poly(A) RNA purified from the insoluble stress-granule-enriched fraction and poly(A) RNA prepared from total NIH3T3 cellular extracts. m^{6}A levels were quantified by thin-layer chromatography, and normalized to the combined intensities of A, C and U. In non-stressed cells, there was no significant difference in the levels of m^{6}A-mRNA in the total cellular or insoluble RNA fraction. By contrast, a significant increase in m^{6}A levels was detected in the stress-granule fraction obtained from either heat-shocked or arsenite-stressed cells (control, $n = 3$; heat shock, $n = 4$; arsenite, $n = 4$; where n represents biological replicates). Bar heights represent mean normalized fold change of m^{6}A/(A+C+U) in poly(A) RNA from stress granules compared with the poly(A) RNA from total cellular RNA (control = 1.076, heat shock = 1.49, arsenite = 1.503). Filled circles and diamonds with lines represent paired biological samples. Error bars represent s.e.m. Paired two-sided Student's t-test was performed on unnormalized m^{6}A/(A+C+U) fractions between control and stress conditions. **b**, A cumulative distribution plot of mRNA enrichment in U2OS arsenite-induced stress granules[23] was plotted for mRNAs classified by the number of annotated m^{6}A peaks[31] per transcript. Transcripts with no m^{6}A peaks (that is, non-methylated) are slightly depleted in stress granules relative to total cellular RNA. However, transcripts that contain two or more m^{6}A peaks show enrichment in stress granules in proportion to the number of m^{6}A sites. **c**, **d**, m^{6}A-containing mRNAs show higher enrichment in stress granules compared to non-methylated mRNAs, as visualized using smFISH (**c**). Two mRNAs that lack annotated m^{6}A sites (*Grk6* and *Polr2a*) were matched with m^{6}A-containing mRNAs of similar length and abundance (*Fignl1* and *Fem1b*, four m^{6}A sites each[32]). *Grk6* and *Polr2a* are not enriched in stress granules (**d**). *Fignl1* and *Fem1b* are markedly more enriched within stress granules as a fraction of total smFISH puncta after heat-shock stress. In **c**, representative slices from confocal Z-stacks are shown to demonstrate localization. In **d**, images (*Grk6/Fignl1*: $n = 5$ images, 26 cells, 2 biological replicates; *Polr2a/Fem1b*: $n = 5$ images, 24 cells, 2 biological replicates) were analysed to assess mRNA localization to stress granules. Bar heights represent mean fraction of stress-granule-localized smFISH puncta and error bars represent s.e.m. Two-sided Student's t-test. **e**, mRNA expression levels were determined by RNA sequencing before and after heat shock (42 °C, 30 min). Transcript abundance was unaltered for non-methylated, singly methylated and polymethylated m^{6}A-mRNAs. **f**, Translation efficiency before heat shock was calculated using matched ribosome-profiling and RNA sequencing data and compared for each mRNA in the methylated state (that is, in wild-type cells) versus the non-methylated state (that is, in *Mettl14*-knockout cells). Transcripts were binned on the basis of the annotated number of m^{6}A sites as in **e**. The centre of the box plot represents the median log$_2$-transformed fold change, the boundaries contain genes within a quartile of the median, and the whiskers represent genes in the upper and lower quartiles. m^{6}A-mRNAs in wild-type mES cells did not show a significant difference in translational efficiency compared to *Mettl14*-knockout mES cells. n denotes the number of genes in each bin. Binned gene groups with annotated m^{6}A sites were compared to genes with no m^{6}A sites with an unpaired two-sided Student's t-test. **g**, Translation efficiency in wild-type and *Mettl14*-knockout mES cells subjected to 30 min of continuous heat shock (42 °C, 30 min) followed by 1 h recovery at 37 °C. Only polymethylated transcripts showed significantly decreased translation efficiency. The effect of m^{6}A was determined by comparing the translation efficiency for each transcript in the methylated form (wild-type cells) relative to the same transcript in the non-methylated form (*Mettl14*-knockout cells). The same binning and m^{6}A annotation strategy were used as in **f**. Boxplots are presented as in **f**. n denotes the number of genes in each bin (total number of genes: 6,720 (**f**) and 7,131 (**g**)). Binned gene groups with annotated m^{6}A sites were compared to genes with no m^{6}A sites with an unpaired two-sided Student's t-test. ****$P < 0.0001$.

m^{6}A can be enriched in stress granules (Fig. 4b), which suggests that mRNAs can also be recruited through m^{6}A-independent mechanisms.

It is possible that phase partitioning of DF–m^{6}A-mRNA complexes to different phase-separated compartments would impart a different fate to m^{6}A-mRNAs. In the case of unstressed cells, the targeting of DF2 and m^{6}A-mRNA to P-bodies facilitates the degradation of m^{6}A-mRNA[25]. We therefore asked whether the relocalization of DF proteins and m^{6}A-mRNA to stress granules affects mRNA abundance. Cellular mRNA levels were examined using RNA sequencing in wild-type mES cells before heat shock, immediately after heat shock, and after a one-hour recovery period. The results showed no substantial alteration in the expression of m^{6}A-modifed mRNAs (Fig. 4e, Extended Data Fig. 5a, Supplementary Table 1), which suggests that DF–m^{6}A-mRNA complexes in stress granules do not induce mRNA degradation.

We used ribosome profiling to compare the translation efficiency of mRNAs when they contain m^{6}A (in wild-type cells) relative to when they lack m^{6}A (in *Mettl14*-knockout cells). In this way, we can determine the effect of m^{6}A on each mRNA in the dataset. Before heat shock, we found no substantial effect of m^{6}A on translation efficiency (Fig. 4f, Supplementary Table 2). As expected, ribosome-protected fragments were markedly reduced after 30 min of heat shock, which is consistent with the global translational suppression reported during most

stresses[26] (Extended Data Fig. 5b). However, translation was detectable one hour after the cessation of heat shock (Extended Data Fig. 5c, d). At this time point, transcripts containing four or more m^{6}A sites showed significantly reduced translational efficiency in wild-type relative to knockout cells (Fig. 4g, Supplementary Table 2). Therefore, polymethylated mRNAs are preferentially repressed, potentially as a result of their phase separation.

Our studies demonstrate that m^{6}A regulates the fate of cytosolic mRNA by scaffolding DF proteins, which leads to the formation of phase-separated DF–m^{6}A-mRNA complexes that then partition into phase-separated structures in cells. This effect is especially efficient for polymethylated mRNAs, which can scaffold multiple DF proteins. Although mRNAs are targeted to diverse intracellular condensates through diverse RNA–RNA and RNA–protein interactions[27], the presence of m^{6}A further enhances the partitioning into these structures. Furthermore, singly methylated and polymethylated mRNAs have different fates, which probably reflect their different abilities to promote LLPS. Notably, monomethylated and polymethylated mRNAs are linked to distinct cellular functions and biological processes (Extended Data Fig. 6a, b). LLPS may therefore influence specific cellular processes by selectively affecting the translation of mRNAs on the basis of their polymethylation status.

Unlike other forms of RNA-scaffolded LLPS, m^{6}A provides a mechanism for regulated phase separation based on the multivalency of m^{6}A. Because m^{6}A levels might vary in different disease, differentiation or signalling contexts[14,28], the phase-separated transcriptome will be encoded, in part, by the cell context-specific distribution and number of m^{6}A sites in each mRNA (Extended Data Fig. 7a, b). The efficiency of m^{6}A-dependent regulation of an mRNA will probably be determined by pathways that control the efficiency of LLPS of the DF proteins.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests are available at https://doi.org/10.1038/s41586-019-1374-1.

Received: 17 January 2019; Accepted: 13 June 2019;
Published online 10 July 2019.

1. Perry, R. P. & Kelley, D. E. Existence of methylated messenger RNA in mouse L cells. *Cell* **1**, 37–42 (1974).
2. Desrosiers, R., Friderici, K. & Rottman, F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc. Natl Acad. Sci. USA* **71**, 3971–3975 (1974).
3. Meyer, K. D. & Jaffrey, S. R. Rethinking m^{6}A readers, writers, and erasers. *Annu. Rev. Cell Dev. Biol.* **33**, 319–342 (2017).
4. Patil, D. P., Pickering, B. F. & Jaffrey, S. R. Reading m^{6}A in the transcriptome: m^{6}A-binding proteins. *Trends Cell Biol.* **28**, 113–127 (2018).
5. Nott, T. J. et al. Phase transition of a disordered nuage protein generates environmentally responsive membraneless organelles. *Mol. Cell* **57**, 936–947 (2015).
6. Lin, Y., Protter, D. S., Rosen, M. K. & Parker, R. Formation and maturation of phase-separated liquid droplets by RNA-binding proteins. *Mol. Cell* **60**, 208–219 (2015).
7. Quiroz, F. G. & Chilkoti, A. Sequence heuristics to encode phase behaviour in intrinsically disordered protein polymers. *Nat. Mater.* **14**, 1164–1171 (2015).
8. Wiśniewski, J. R., Hein, M. Y., Cox, J. & Mann, M. A "proteomic ruler" for protein copy number and concentration estimation without spike-in standards. *Mol. Cell. Proteomics* **13**, 3497–3506 (2014).
9. Shin, Y. & Brangwynne, C. P. Liquid phase condensation in cell physiology and disease. *Science* **357**, eaaf4382(2017).
10. Molliex, A. et al. Phase separation by low complexity domains promotes stress granule assembly and drives pathological fibrillization. *Cell* **163**, 123–133 (2015).
11. Kroschwald, S. et al. Promiscuous interactions and protein disaggregases determine the material state of stress-inducible RNP granules. *eLife* **4**, e06807 (2015).
12. Standart, N. & Weil, D. P-bodies: cytosolic droplets for coordinated mRNA storage. *Trends Genet.* **34**, 612–626 (2018).
13. Zhou, J. et al. Dynamic m^{6}A mRNA methylation directs translational control of heat shock response. *Nature* **526**, 591–594 (2015).
14. Dominissini, D. et al. Topology of the human and mouse m^{6}A RNA methylomes revealed by m^{6}A-seq. *Nature* **485**, 201–206 (2012).
15. Meyer, K. D. et al. 5′ UTR m^{6}A promotes cap-independent translation. *Cell* **163**, 999–1010 (2015).
16. Anders, M. et al. Dynamic m^{6}A methylation facilitates mRNA triaging to stress granules. *Life Sci. Alliance* **1**, e201800113 (2018).
17. Slobodin, B., Han, R., Calderone, V., Vrielink, J. A., Loayza-Puch, F., Elkon, R. & Agami, R. Transcription impacts the efficiency of mRNA translation via co-transcriptional *N*^{6}-adenosine methylation. *Cell* **169**, 326–337.e312 (2017).
18. Knuckles, P. et al. RNA fate determination through cotranscriptional adenosine methylation and microprocessor binding. *Nat. Struct. Mol. Biol.* **24**, 561–569 (2017).
19. Wang, P., Doxtader, K. A. & Nam, Y. Structural basis for cooperative function of Mettl3 and Mettl14 methyltransferases. *Mol. Cell* **63**, 306–317 (2016).
20. Wang, X. et al. Structural basis of *N*^{6}-adenosine methylation by the METTL3–METTL14 complex. *Nature* **534**, 575–578 (2016).
21. Śledź, P. & Jinek, M. Structural insights into the molecular mechanism of the m^{6}A writer complex. *eLife* **5**, e18434 (2016).
22. Han, T. W. et al. Cell-free formation of RNA granules: bound RNAs identify features and components of cellular assemblies. *Cell* **149**, 768–779 (2012).
23. Khong, A., Matheny, T., Jain, S., Mitchell, S. F., Wheeler, J. R. & Parker, R. The stress granule transcriptome reveals principles of mRNA accumulation in stress granules. *Mol. Cell* **68**, 808–820.e805 (2017).
24. Namkoong, S., Ho, A., Woo, Y. M., Kwak, H. & Lee, J. H. Systematic characterization of stress-induced RNA granulation. *Mol. Cell* **70**, 175–187. e178 (2018).
25. Wang, X. et al. *N*^{6}-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505**, 117–120 (2014).
26. Bouche, G., Amalric, F., Caizergues-Ferrer, M. & Zalta, J. P. Effects of heat shock on gene expression and subcellular protein distribution in Chinese hamster ovary cells. *Nucleic Acids Res.* **7**, 1739–1747 (1979).
27. Van Treeck, B. & Parker, R. Emerging roles for intermolecular RNA–RNA interactions in RNP assemblies. *Cell* **174**, 791–802 (2018).
28. Meyer, K. D. et al. Comprehensive analysis of mRNA methylation reveals enrichment in 3′ UTRs and near stop codons. *Cell* **149**, 1635–1646 (2012).
29. Kedersha, N. et al. Stress granules and processing bodies are dynamically linked sites of mRNP remodeling. *J. Cell Biol.* **169**, 871–884 (2005).
30. Zhu, T. et al. Crystal structure of the YTH domain of YTHDF2 reveals mechanism for recognition of *N*^{6}-methyladenosine. *Cell Res.* **24**, 1493–1496 (2014).
31. Xiang, Y. et al. RNA m^{6}A methylation regulates the ultraviolet-induced DNA damage response. *Nature* **543**, 573–576 (2017).
32. Geula, S. et al. m^{6}A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation. *Science* **347**, 1002–1006 (2015).

METHODS

Cell culture. HEK 293T/17 (ATCC CRL-11268) cells, U2OS (ATCC HTB-96) cells, and NIH3T3 (ATCC CRL-1658) cells were obtained from ATCC. Cells were maintained in DMEM (11995-065, Thermo Fisher Scientific) with 10% FBS, 100 U ml^{-1} penicillin and 100 μg ml^{-1} of streptomycin under standard tissue-culture conditions (37 °C, 5.0% CO_2). Mycoplasma contamination in cells was routinely tested by Hoechst staining. *Mettl14*-knockout and wild-type mouse ES cells have been previously described[19] and were a gift from J. Hanna and S. Geula (Weizmann Institute of Science). ES cells were grown in Knockout DMEM (10829018, Invitrogen) with 15% heat-inactivated FBS, 100 U ml^{-1} penicillin and 100 μg ml^{-1} of streptomycin, 200 mM L-glutamine, 1% non-essential amino acids (Gibco 11140076), 50 μM β-mercaptoethanol (21985023, Gibco), 1,000 U ml^{-1} mouse LIF (ESG1107, EMD Millipore), 3 μM GSK3 inhibitor CHIR99021 (04-0004-02, Stemgent) and 1 μM MEK1 inhibitor PD0325901 (04-0006-02, Stemgent). Media was changed daily. Cells were cultured on 0.1% gelatin-coated (07903, StemCell Technologies) plates and grown under standard tissue-culture conditions (37 °C, 5.0% CO_2). Cells were passaged as needed using TrypLE Express (Life Technologies) according to the manufacturer's instructions.

Stress conditions were induced as follows: heat shock at 42 °C for 30 min in a water bath; arsenite stress with 0.5 mM of sodium arsenite (35000-1L-R, Fluka) for 30–60 min, as indicated; and thapsigargin stress with 10 μM for 3 h (mES cells) or 5 μM for 1.5 h (NIH3T3 cells). All stress experiments were performed in duplicate or triplicate. Investigators were not sample-blinded, and no randomization of samples was performed.

Immunostaining. Cells were plated to reach 40–60% of confluency the following day on a 35-mm Petri dish coated with poly-D-lysine (PG35GC-1.5-14-C). For mES cell culture and immunostaining, Petri dishes were coated with gelatin for 1 h at 37 °C. Cells were washed three times with PBS and fixed for 15 min with 4% paraformaldehyde in PBS. Cells were permeabilized and blocked with 0.2% Triton X-100 and 2% FBS in PBS for 30 min at 25 °C. Cells were incubated for 90 min with the primary antibody followed by washing three times in PBS. After washing, cells were incubated with secondary antibodies conjugated to Alexa Fluor 488 and/or Alexa Fluor 594 at 2 μg ml^{-1} in PBS (Life Technologies) for 60 min. Nuclei were stained with Hoechst 33342 (66249, Life Technologies) at a 0.1 μg ml^{-1} in PBS for 10 min. Coverslips were mounted using Prolong Diamond Antifade Mountant (P36961, Life Technologies). All immunostaining steps were carried out at 25 °C.

DF2 localization in P-bodies was determined using the P-body marker EDC4. The large number of P-bodies seen in the *Mettl14*-knockout cells compared to wild-type cells is due to the different morphology of *Mettl14*-knockout cells. As has been described previously for *Mettl14*-knockout and other m^{6}A-deficient cells, m^{6}A-depleted cells are flattened, whereas wild-type cells are 'dome-shaped'[33,34]. As a result, in a single confocal slice, more P-bodies are seen in *Mettl14*-knockout cells. By contrast, because wild-type cells are dome shaped, there are many more *z*-stacks, and the P-bodies are found throughout the different confocal slices. However, overall, there is no substantial difference in the number of P-bodies in wild-type and *Mettl14*-knockout cells.

The puromycin time course experiment was performed as follows: cells were heat-shocked at 42 °C in a water bath for 30 min and incubated with 10 μg ml^{-1} of puromycin before each time point for 10 min and then washed with PBS and fixed with 4% paraformaldehyde. Staining was performed using anti-TIAR (5137S Cell Signalling Technology) and anti-puromycin (NC0327811, Millipore Sigma) antibodies.

Protein expression and purification. N-terminal 6×-His tagged DF1, DF2, DF3 and YTH domain were generated by a standard PCR-based cloning strategy from HEK293T oligo-d(T)25-primed cDNA as described previously[35]. DF proteins and the YTH domain were overexpressed in *Escherichia coli* Rosetta2 (DE3) single (Novagen) using pET-28(+) (Novagen) or pProEx HTb (Invitrogen). *E. coli* expressing DF proteins and the YTH domain were induced with 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) for 16 h at 18 °C. Cells were collected, pelleted and then resuspended in the following buffer: 50 mM NaH_2PO_4 pH 7.2, 300 mM NaCl, 20 mM imidazole at pH 7.2 and supplemented with EDTA-free protease inhibitor cocktail (05892791001, Roche) according to the manufacturer's instructions. The cells were lysed by sonication and then centrifuged at 10,000*g* for 20 min. The soluble protein was purified using Talon Metal Affinity Resin (Clontech) and eluted in the following buffer: 50 mM NaH_2PO_4 pH 7.2, 300 mM NaCl, 250 mM imidazole-HCl at pH 7.2. Further concentration and buffer exchange was performed using Amicon Ultra-4 spin columns (Merck-Millipore). Recombinant protein was stored in the following buffer: 20 mM HEPES pH 7.4, 300 mM KCl, 6 mM $MgCl_2$, 0.02% NP40, 50% glycerol at −80 °C or 20% glycerol at −20 °C. All protein purification steps were performed at 4 °C. The purified protein was quantified using a ND-2000C NanoDrop spectrophotometer (NanoDrop Technologies) with OD 280 and verified by Coomassie staining.

Protein labelling. For DF phase-separation experiments, DF1, DF2 and DF3 proteins were fluorescently labelled using Alexa Fluor (488, 594 and 647, respectively) Microscale Protein Labelling kit according to the manufacturer's instructions (A30006, A30008, A3009, Thermo Fisher Scientific). In brief, DF proteins were diluted at 1 mg ml^{-1} in PBS and mixed with 100 mM sodium bicarbonate. The reaction was incubated for 15 min at room temperature and fluorescently labelled proteins were purified from the unreacted dye substrate by column purification using Micro Bio-Spin Columns with P-30 gel. The labelled protein was eluted in 20 mM HEPES pH 7.4, 300 mM KCl, 6 mM $MgCl_2$ and 0.02% NP-40 and buffer exchange was performed in two successive rounds using Amicon 0.5 ml Ultracel centrifugal filter columns. Protein labelling was performed on the day of each experiment.

Droplet formation. DF2 was purified as described previously[36]. Temperature-dependent droplet assembly was performed in the following buffer: 20 mM HEPES pH 7.4, 300 mM KCl, 6 mM $MgCl_2$, 0.02% NP-40. For non-fluorescent DF2 (75 μM), droplet-containing buffer was placed on a coverslip and visualized by either phase-contrast or differential interference contrast microscopy using a Nikon TE-2000 inverted microscope. Temperature-dependent phase separation experiments were performed by incubating DF2 at 37 °C for 1 min after removal from ice. The temperature-dependent phase-transition diagram was generated by visualizing droplets using phase-contrast microscopy on a coverslip incubated in a temperature-, humidity- and CO_2-controlled top stage incubator (Tokai Hit). The temperature was increased from 22 °C to 37 °C at a rate of 1 °C per minute and images were taken every 30 s.

RNA-dependent droplet-formation experiments were performed in the following buffer: 10 mM HEPES pH 7.4, 150 mM KCl, 3 mM $MgCl_2$, 0.01% NP-40 and 10% glycerol. DF2 (25 μM) diluted in buffer was placed on a coverslip and RNA containing 0, 1 or 10 m^{6}A nucleotides was added (570 nM). The solution was incubated at 37 °C for 10 min and droplets were visualized with phase-contrast microscopy.

The salt-dependent phase separation was generated by combining diluted DF2 protein (1–8 μM) with NaCl buffer (20 mM HEPES pH 7.4, 300 mM KCl, 6 mM $MgCl_2$, 0.02% NP-40, 50% glycerol, with NaCl) on a coverslip and scoring yes/no for the presence of protein droplets as previously described[37] by observation using a bright-field microscope.

In vitro transcription. To synthesize RNAs containing a single m^{6}A or A nucleotide, or 10 m^{6}A or 10 A nucleotides, we performed in vitro transcription using reactions that contained either m^{6}A triphosphate or ATP. This approach ensures that all adenosines are either in the m^{6}A or the A form. In vitro transcription was performed using AmpliScribe T7 High Yield Transcription kit (AS3107, Lucigen) according to the manufacturer's instructions. The template encodes an RNA containing a single adenosine (indicated in bold): (GGTCTCGGTCTTGGTCTCTGGTCTTTGG**A**CTTGGTCT TGGTCTTCG GTCTCGGTC TTTGGTCT) or 10 adenosines in the canonical GGACU consensus motif for m^{6}A: (GG**A**CTCGG**A**CTTGG**A**CTCTGG**A**CTTTGG**A**CTT GG**A**CTTGG**A**CTTCGG**A**CTCGG**A**CTTTGG**A**CT). The m^{6}A versions of the RNA were synthesized by replacing adenosine 5′ triphosphate in the reaction by N^6-methyadenosine 5′ triphosphate (TriLink). The reaction was terminated by the addition of DNase I and incubation for 15 min at 37 °C. RNA was purified using an Oligo Clean and Concentrator column (D4061, Zymo Research). RNA concentration was determined using a NanoDrop spectrophotometer and verified by TBE-urea denaturing gel electrophoresis. Nucleic acid staining was performed with SYBR Gold (S11494). DNA matrix was obtained by hybridizing DNA oligonucleotides containing a T7 promoter and the target sequence.

For fluorescent RNA in vitro transcription, BODIPY FL-guanosine 5′-*O*-(3-thiotriphosphate) fluorescent GTPs (G22183, Invitrogen) were added to the reaction in a 1:10 molar ratio with GTPs. The thiotriphosphate linkage prohibits the fluorescent nucleotides from being internally incorporated, and only allows incorporation at the +1 position of in vitro transcripts (the initial G after the T7 promoter sequence). Incorporation of the fluorescent GTP into transcripts was verified by TBE-urea denaturing gel electrophoresis and fluorophore excitation by exposure to 488-nm light. RNA concentrations were determined using a NanoDrop spectrophotometer and verified by SYBR Gold staining.

Partition coefficients. For partition coefficient experiments with Alexa488-labelled DF proteins, DF proteins (15 μM) were mixed in a buffer containing 20 mM HEPES 7.6, 300 mM KCl, 6 mM $MgCl_2$, 0.02% NP-40 and 50% glycerol. Upon addition of 425 nM RNA containing 10 m^{6}A nucleotides, the reaction was held at 37 °C for approximately 10 min. DF-containing droplets were then imaged at 40× using a bright-field microscope. Partition coefficients for the no-RNA condition were calculated by creating a ratio of DF intensity in solution over DF intensity located in the immediately adjacent region. After DF2 droplet enrichment following the addition of m^{6}A-RNA, partition coefficients were calculated for stably formed DF-containing droplets. Follow-up partition coefficient calculations

were performed approximately 24 h after the initial formation of DF2-enriched droplets.

For fluorescent m^{6}A-RNA partition coefficients, fluorescent RNAs (850 nM) were mixed with a 7.5 μM solution of DF2 in buffer containing 10 mM HEPES 7.6, 150 mM KCl, 3 mM $MgCl_2$, 0.01% NP-40 and 10% glycerol. Formation of fluorescent DF2:m^{6}A-RNA coacervates was visualized by fluorescence microscopy within minutes. Partition coefficients were calculated by taking the ratio of fluorescence intensity of soluble fluorescent m^{6}A-RNA over m^{6}A-RNA-enriched DF2 droplets and adjacent regions.

For in vivo DF2 intensity ratios in mES cells, TIAR staining was used to demarcate stress granule boundaries. Regions of interest (ROIs) were manually drawn to encompass a central portion of the TIAR-positive stress granules and the fluorescence intensity values for DF2 in these regions were averaged. Intensity ratios of DF2 staining immediately adjacent to the TIAR-positive stress granules in an identical ROI were used as background. Intensity ratios were then determined by calculating the average phase-separated DF2 intensity inside stress granules over the average soluble DF2 intensity in the immediately adjacent cytoplasm.

Determination of relative m^{6}A levels by two-dimensional thin layer chromatography. Experiments measuring m^{6}A in stress granules and in the cytosol were performed and analysed by an investigator blinded to the sample identity. Relative levels of internal m^{6}A in mRNA were determined by thin-layer chromatography (TLC) as described previously[38]. This method selectively examines m^{6}A in the mRNA sequence context, thereby preventing problems with contamination of the m^{6}A signal by co-purifying ribosomal RNA or small nuclear RNA. Additionally, levels of adenosine in the poly(A) tail are not measured because only nucleotides (methylated or non-methylated) that are followed by G are detected in this assay. In brief, twice-purified poly(A) RNA was digested with 2 U of RNase T1 (Thermo Fisher Scientific) for 2 h at 37 °C in the presence of RNaseOUT (Invitrogen). Digested 5′ ends RNA were subsequently labelled with 10 units of T4 PNK (New England Biolabs) and 0.4 mBq [γ-^{32}P]ATP for 30 min at 37 °C followed by removal of the γ-phosphate of ATP by incubation with 10 U apyrase (New England Biolabs) at 30 °C for 30 min. After phenol–chloroform extraction and ethanol precipitation, RNA samples were resuspended in 10 μl of H_2O and digested to single nucleotides with 2 units of P1 nuclease (Sigma) for 3 h at 37 °C. The released 5′ monophosphates from this digest (1 μl) were then analysed by 2D TLC on glass-backed PEI-cellulose plates (MerckMillipore) as previously described[38]. No m^{6}A was detected in *Mettl14*-knockout poly(A) RNA, which is consistent with mass spectrometry data that have previously been obtained from this cell line[19].

Synthesis and cloning of mNeonGreen open reading frame. Monomeric NeonGreen (mNeonGreen) protein-coding open reading frame (ORF) was synthesized in vitro using overlapping 60-mer DNA oligonucleotides designed using DNAWorks[39]. In brief, a HindIII restriction-site-deficient, human codon-optimized DNA sequence for mNeonGreen protein sequence (GenBank: AGG56535.1) and 60-mer overlapping DNA oligos were generated using DNAWorks. The ORF was synthesized in two PCR reaction steps. In the first PCR, oligo assembly PCR amplification was performed by mixing the overlapping oligos at 2 μM in 1X Phusion HF master mix (New England Biolabs, cat. no. M0531S) and PCR cycling at 98 °C for 30 s; 25 cycles of 98 °C for 5 s, 64 °C for 20 s, 72 °C for 20 s; 72 °C for 10 s; 4 °C hold. Oligo assembly PCR mixture (1 μl) was subjected to a second PCR in which the open reading frame was amplified using mNeonGreen forward (ATATAAGCTTGATATGGTGAGTAAGGGCGAAGAGGA) and reverse (ATATAAGCTTTTTATACAACTCGTCCATGCCCATCACG) primers in 50 μl of 1X Phusion HF master mix and the following thermal cycling conditions: 98 °C for 30 s; 30 cycles of 98 °C for 5 s, 64 °C for 20 s, 72 °C for 20 s; 72 °C for 10 s; 4 °C hold. The amplified PCR product (of 731 bp) was gel-eluted, digested with HindIII, and cloned at HindIII site at pcDNA4/TO Mammalian Expression Vector (Thermo Fisher Scientific, cat. no. V102020). Bacterial clones containing mNeonGreen ORF in the correct orientation were selected by DNA sequencing. This plasmid is referred to as pcDNA-4TO-mNeonGreen.

Cloning and generation of DF2 and DF2-mutant plasmids. Human DF2 was amplified by PCR using DF2-BamHI-F (ATATGGATCCATGTCGGCCA GCAGCCTCTT) and DF2-XhoI-R (GGTGCTCGAGCTATTTCCCACGACCT TGACGTTCCTT) oligonucleotides using human cDNA made by oligo dT-priming HEK-293T total RNA. The PCR product was gel eluted and digested and cloned at BamHI and XhoI in pcDNA-4TO-mNeonGreen plasmid. The W432A mutation was introduced using DF2-W432A-SDM-F (GCGTGCAGCACAGAGCATGG) and DF2-W432A-SDM-R (AATATTATACTTAATGGAACGGTGAATATCGTCC) oligonucleotides in 1X Phusion HF master mix.

CRISPR–Cas9 knock-in of NeonGreen into the endogenous *YTHDF2* locus. For fluorescence recovery after photobleaching (FRAP) experiments of DF2 in stress granules, NeonGreen was inserted into the endogenous locus using CRISPR, because plasmid-based expression of DF2 in HEK293 cells is associated with the formation of ectopic granules. Knock-in by CRISPR was performed as described previously[40]. The sequence of the guide RNA used is (TGTAGGAACGTCAAGGTCGT). For these experiments we generated a single-stranded homology directed repair DNA template containing 800-nucleotide-long homology arms flanking a NeonGreen coding sequence immediately before the stop codon of DF2. Successful incorporation was validated by western blotting using a DF2-specific antibody for NeonGreen–DF2, which exhibited the expected mobility shift relative to DF2.

Antibodies. The following antibodies were used for immunofluorescence experiments: rabbit anti-TIAR (5137S, Cell Signaling Technology, lot no. 1, 1:100), mouse anti-TIAR (Clone 6) (610352, BD Biosciences, lot numbers 5357680; 7219778, 1:100), mouse anti-Edc4 (H-12) (sc-376382, Santa Cruz Biotechnology, lot no. I0216, 1:100), mouse anti-Puromycin clone 12D10 (MABE343, Millipore Sigma, lot no. 2861354, 1:100), rabbit anti-G3BP1 (13057-2-AP, Proteintech, lot no. 00047654, 1:100), mouse anti-ATXN2 (Clone 22) (611378, BD Biosciences, lot no. 7341666, 1:100), rabbit anti-YTHDF1 (17479-AP, Proteintech, lot no. 00040713, 1:100), rabbit anti-YTHDF2 (24744-1-AP, Proteintech, lot no. 00053880, 1:100), rabbit anti-YTHDF3 (ab103328, Abcam, lot no. GR35115-39, 1:100), rabbit anti-IgG Alexa Fluor 594 (A11012, Invitrogen, lot no. 1933366, 1:1,000), mouse anti-IgG Alexa Fluor 488 (A11001, Invitrogen, lot no. 1939600, 1:1,000).

The following antibodies were used for immunoblotting experiments: rabbit anti-YTHDF2 (ARP67917_P050, Aviva System Biology, lot no. QC38405-43182, 1:1,000), mouse anti-GAPDH (SC-365062, Santa Cruz Biotechnology, lot no. A2816, 1:5,000), rabbit anti-IgG HRP (NA934V, GE Healthcare, lot no. 16677077, 1:10,000), mouse anti-IgG HRP (NA931V, GE Healthcare, lot no. 16814909, 1:10,000).

Overexpression of DF2 in mES cells. Wild-type and *Mettl14*-knockout mES cells were transfected with NeonGreen-tagged DF2- and DF2(W432A)-expressing plasmids using Fugene HD transfection Reagent (E2311, Promega) according to the manufacturer's instructions. In brief, cells were plated on 35-mm glass-bottom dishes coated with poly-D-lysine (PG35GC-1.5-14-C) and 1% gelatin and allowed to reach 40–60% confluency the following day. 48 h after transfection, plates were placed in a temperature, humidity and CO_2-controlled stage-top incubator for live-cell imaging (Tokai Hit) and cells were imaged by fluorescence microscopy. Notably, the expression of DF2 in mES cells was not associated with the ectopic formation of granules, as can be seen in the images. This contrasts with HEK293 cells and other cell types, in which we found that DF2 expression caused the formation of granules. We therefore performed these experiments in mES cells. mES cells were then subject to heat shock at 42 °C in a water bath for 30 min. Cells were then washed twice with PBS, fixed for 15 min with 4% paraformaldehyde, and again washed twice with PBS before visualizing by fluorescence microscopy.

Single-molecule FISH. Cells were plated to reach 40–60% confluency the following day on a 35 mm Petri dish coated with poly-D-lysine. For mES cell culture and immunostaining, Petri dishes were coated with gelatin for 1 h at 37 °C. Single-molecule fluorescence of mRNA was performed using the ViewRNA Cell Plus Assay kit (Invitrogen, 88-19000). All steps were carried out according to the manufacturer's instructions. Alexa594-labelled and Alexa647-labelled ViewRNA probes (Invitrogen) were used to detect the presence of non-methylated and polymethylated mRNAs in the same cell sample. TIAR was stained as a stress granule marker. Wild-type mES cells ($n = 15$) were analysed by confocal microscopy.

For smFISH, we compared the non-methylated mRNAs *Grk6* (length = 2,994 nucleotides, normalized RNA sequencing (RNA-seq) counts = 712) and *Polr2a* (length = 6,740 nucleotides, normalized RNA-seq counts = 2,963), and compared them with matched m^{6}A-containing mRNAs *Fignl1* (length = 2,974 nucleotides, normalized RNA-seq counts = 892; 4 annotated m^{6}A peaks[19]) and *Fem1b* (length = 6,785 nucleotides, normalized RNA-seq counts = 2,396; 4 annotated m^{6}A peaks[19]). Probes were labelled with different fluorophores so that comparisons of stress-granule localization could be performed in the same cells for each matched probe pair.

Western blotting. Cells were lysed in whole-cell lysis buffer (10 mM Tris-HCl pH 7.4, 10 mM EDTA, 50 mM NaF, 50 mM NaCl, 1% Triton X-100, 0.1% SDS, with 1X protease and phosphatase inhibitor (78440, Pierce)) and sonicated. Protein quantification was performed using the Pierce BCA protein assay kit according to the manufacturer's instructions (23227, Thermo Fisher Scientific). Equal quantities of proteins were separated on 4–12% Bis–Tris gels (Invitrogen) and transferred onto a nitrocellulose membrane for 1 h at a constant voltage of 30 V at 4 °C. Membranes were blocked by incubation in 5% milk in TBS-T for 1 h at room temperature under agitation. Membranes were stained with primary antibodies, extensively washed in TBS-T and then incubated with appropriate secondary antibodies conjugated to HRP. Blots were imaged on a ChemiDoc XRS+ system (Bio-Rad).

RNA-seq analysis. mES cells were plated on a 10-cm dish and allowed to reach 70–80% confluence. Heat shock was performed for 30 min in a water bath at 42 °C. To measure RNA expression after stress, heat-shocked cells were placed back at 37 °C for 1 h. After collecting cells as described in the section 'Ribosome profiling', 5% of the lysate cells before RNase digestion was used to extract total RNA. Library

preparation was performed using the NEBNext Ultra Directional RNA Library Prep Kit starting from 1 μg of total input RNA and following the protocol for use with NEBNext rRNA Depletion Kit. The libraries were sequenced on the Illumina HiSeq 2500 instrument, in single-read mode, with 50 bases per read. A separate independent biological replicate was sequenced so that each ribosome profiling (Ribo-seq) replicate had an RNA-seq partner to be paired with for translational efficiency analysis.

After sequencing, fastq files were trimmed for quality and read lengths shorter than 16 nucleotides were discarded. The adaptor was removed using FLEXBAR[41]. Duplicates were removed with the pyFastDuplicateRemover.py utility from the PyCRAC software suite as previously described[41]. Ribosomal reads were removed using STAR aligner[42]. The remaining reads were mapped to the mm10 genome using STAR and the data were used to normalize the Ribo-seq data for the translational efficiency measurements. A pseudocount of 0.001 was added to avoid division by zero. Differential expression analysis of RNA-seq data was performed with the DESeq2 package in R. Differential expression data used to generate the plots for Fig. 4e and Extended Data Fig. 5a are available in Supplementary Table 1. Raw data are available at NCBI Gene Expression Omnibus (GEO) under accession number GSE125725.

Ribosome profiling. Ribosome profiling was performed essentially as previously described[43]. In brief, mES cells were plated on a 10-cm dish and allowed to reach 70–80% confluence. Heat shock was performed for 30 min in a water bath at 42 °C. To measure translation after stress, heat-shocked cells were placed back at 37 °C for 1 h to enable translation to resume. To inhibit ribosome transit post-lysis, cells were rapidly washed twice with ice-cold PBS containing 50 μg ml^{-1} of cycloheximide. To generate ribosome-protected fragments, cells were pelleted and immediately lysed in 400 μl of cell lysis buffer (20 mM Tris pH 7.4, 150 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT, 100 μg ml^{-1} cycloheximide, 25 U DNase I). Lysate was clarified by performing a centrifugation step at 20,000*g* for 10 min at 4 °C. Supernatant was collected. A 5% fraction of this supernatant was used for RNA-seq preparation. RNA (30 μg) was digested with RNase I to isolate ribosome-protected fragments. After RNase digestion, lysates were loaded on a sucrose gradient and centrifuged in a TLA-100.3 fixed-angle rotor at 100,000 r.p.m. to recover ribosome-protected fragments. RNA from the resuspended ribosomal pellet was purified and run on a gel to selectively excise footprinted RNAs (from 17- to 34-nucleotide RNA fragments).

To avoid ribosomal contamination in the library preparation steps, we then performed Ribo-Zero Gold depletion of the footprinted RNA. The ribosomal-RNA-depleted RNA fragments were dephosphorylated and the linker was added. To specifically deplete unligated linker, yeast 5′-deadenylase and RecJ exonuclease digestion was performed. At this point, the library preparation steps were performed essentially as previously[41]. In brief, reverse transcription was performed using SuperScript III. To avoid untemplated nucleotide addition, reverse transcription was carried out at 57 °C, as previously described[43]. cDNA purification, circularization and amplification were performed as previously described[41]. Libraries were sequenced with a single-end 50-bp run using an Illumina Hiseq2500 platform. Generation of the sequencing libraries was from four separately purified biological replicates. After sequencing, fastq files were quality-based trimmed and reads below 16 nucleotides were excluded. The adaptor was removed using the FLEXBAR tool[41]. The demultiplexing was performed on the basis of the experimental barcode using the pyBarcodeFilter.py script[41]. The second part of the iCLIP random barcode was then moved to the read header (awk -F "##" '{sub(/..../,"##"$2, $2); getline($3); $4 = substr($3,1,2); $5 = substr($3,3); print $1 $2 $4"\n"$5}').

After removal of the PCR duplicates, ribosomal and mitochondrial RNA reads were removed using STAR aligner[42]. Reads with no acceptable alignment to ribosomal and mitochondrial RNA were then mapped to the mm10 transcriptome. We specifically considered only ribosome-protected fragment reads that mapped to the coding sequence to avoid any possible contamination coming from the untranslated area of the genome. Given our interest in studying translation independently from the initiation and stopping rates, we excluded ribosome-protected fragments mapping to the first 15 amino acids and the last 5 amino acids of each coding sequence. Only the longest splice isoform of each gene was considered. Gene-count tables for ribosome-protected fragments and RNA-sequencing reads from each sample were then normalized and processed using the xtail package in R to calculate translational efficiency[44]. To remove the background, genes with fewer than 4 minimum mean ribosome-protected fragment reads were excluded. Replicates that included less than 50% mapped coding sequence reads were excluded from the final analysis. *Mettl14*-knockout cells were used as the normalizing condition. Translational efficiency tables used to generate the plots in Fig. 4f, g are available in Supplementary Table 2. Raw data can be accessed at the NCBI GEO under accession number GSE125725. Each gene with an assigned log_2-transformed fold change was annotated for the presence of an m^6A site using a previous m^6A mapping study[19].

Analysis of enrichment of methylated RNAs in stress granules. For U2OS cells, stress granule gene expression data (GSE99304) and m^6A methylated RNA immunoprecipitation sequencing (MeRIP-seq) data (GSE92867) were downloaded from the NCBI GEO. For NIH3T3 cells, stress-granule gene-expression data (GSE90869) and m^6A MeRIP-seq data (GSE61998) were downloaded from the GEO. The m^6A bed file was processed to produce a table of m^6A peak counts per gene. The gene expression data was extracted from the fragments per kilobase of transcript per million mapped reads (FPKM) columns for the stress granule and total cell fractions. A pseudocount of 0.001 was added to the expression values to avoid division by zero. The enrichment score was calculated as log_2(stress granule FPKM/total cell FPKM). The cumulative distribution function was calculated for genes grouped by m^6A count and plotted using R.

The abundance of DF proteins stated in the average cell is calculated from an analysis of absolute protein abundance in different cell lines using a proteomic approach[8]. On the basis of a reported number of approximately 740,000 copies of DFs per PC3 cancer cell and an average cytosolic volume of around 2,300 μm^3, we estimate the concentration of DFs to be approximately 5.3 μM in the cell.

FRAP analysis. For in vitro FRAP analysis, fluorescent DF2 droplets were loaded into a cell counter slide (C-Chip DHSC-N01 iN Cyto) at room temperature. The droplet was photobleached in three regions ROIs that were defined for these experiments. ROI-1 was the indicated circular region in the droplet, and ROI-2 was a similarly sized circular region in the same droplet but in an area that was not photobleached. ROI-3 was defined as background and drawn outside the droplet and its signal was subtracted from both ROI-1 and ROI-2. Raw data were plotted using Prism software.

For FRAP experiments of the stress granules in living cells, the entire stress granule was chosen as the ROI in order to more accurately quantify the ability of NeonGreen–DF2 to undergo phase separation from the cytoplasm into the stress granule. Unlike the in vitro experiments above, which involved DF2 droplets that could reach sizes of 10–20 μm in diameter, stress granules in vivo are less than 1 μm in diameter and are mobile. Thus, rather than photobleaching the centre, the entire stress granule was photobleached. Data were normalized to the frame with the highest average ROI intensity level. Bleached granules were subjected to a 514-nm laser burst for 1.03 μs at frame 0. Each frame taken after bleaching represents 3.5 s of recovery. Each data point is representative of the mean and standard deviation of fluorescence intensities in three unbleached (control) or three bleached (experimental) granules.

Gene ontology of U2OS cells. Gene ontology analysis was performed using the PANTHER Gene Enrichment Analysis tool at[45] http://www.geneontology.org. U2OS RNA-seq counts from GSE99304[24] and m^6A MeRIP-seq data from GSE92867[32] were used to generate the input data. Genes that lacked any annotated m^6A sites were classified as having zero sites. For the singly methylated GO, genes with one mapped m^6A site were compared to genes with zero mapped m^6A sites. For the polymethylated GO, genes with four or more mapped m^6A sites were compared to genes with zero mapped m^6A sites. *P* values were calculated using Fisher's exact test with a Bonferroni correction for multiple hypothesis testing. The top twelve genes by *P* value are charted for each gene-ontology category. The minimum *P* value for inclusion was $P < 0.01$.

Protein disorder propensity plots for YTHDF proteins. Protein disorder propensity plots for the YTHDF proteins were prepared using the PLAAC (prion-like amino acid composition) webtool with background set to 0%[46]. Determination of the amino acid composition by per cent was performed using the ProtParam tool from ExPASy (https://web.expasy.org/protparam/) and amino acid composition bar charts were made using ggplot2 in R.

Image acquisition and analysis. Fluorescence imaging and bright-field imaging experiments were performed using a wide-field fluorescent microscope (Eclipse TE2000-E microscope, Nikon). Images were analysed using NIS-Elements Viewer software (Nikon) and Fiji (ImageJ v1.51n) for quantification analysis.

FRAP experiments were performed using an LSM 880 laser scanning confocal microscope (Zeiss) with an Airyscan high-resolution detector connected to a temperature-, humidity- and CO_2-controlled top stage incubator for live-cell imaging (Tokai Hit). Differential interference contrast images were taken with a Zeiss Axioplan 2 upright microscope.

smFISH and P-body experiments were performed using an LSM 880 laser scanning confocal microscope (Zeiss) Airyscan high-resolution detector. *Z*-stacks were taken at 63× oil immersion objective. Images were analysed using ZEN Black software (Zeiss) and Fiji (ImageJ v1.51n). Co-localization and 3D analysis of confocal *Z*-stacks for smFISH experiments were performed using the DiAna plugin for ImageJ[47]. Granules from 5 images with 3–5 cells per image were analysed using this high-throughput method, which enabled us to simultaneously measure smFISH and TIAR antibody signal co-localization for as many as 100 smFISH puncta in a single confocal *Z*-stack. Using this method, the total number of data points from the images for each smFISH probe were scored as a ratio of puncta co-localizing with TIAR-containing stress granules over the fraction of total puncta detected in the cell.

Statistics and reproducibility. All statistical analyses were performed in GraphPad Prism 8 or Microsoft Excel unless otherwise indicated. The outcomes of all statistical tests including *P* values and number of samples are included in the figure panels or corresponding figure legends. Significance was defined as any statistical outcome that resulted in a *P* value of <0.05, unless otherwise indicated. *P* value significance is represented by the following: $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$. Multiple hypothesis correction and *P*-value adjustments were not performed unless otherwise indicated.

Temperature-dependent DF phase separation experiments (Fig. 1a, b) were performed in duplicate on biological replicates. Salt-dependent DF2 phase separation experiments (Fig. 1c) were performed in duplicate on technical replicates. Droplet-formation assays with Alexa488-labelled DF proteins (Fig. 1d, Extended Data Fig. 1b–e) were performed in duplicate on biological replicates. Fluorescence recovery after photobleaching of Alexa488–DF2 was performed on eight droplets in technical replicates. Validation of m^{6}A RNA (Extended Data Fig. 1f) was performed in duplicate from technical replicates. m^{6}A RNA-dependent phase separation experiments (Fig. 1f, g, Extended Data Fig. 1g, h) were performed in duplicate on biological replicates. Staining of DF2 and stress-granule markers in mES cells (Fig. 2a, b, Extended Data Fig. 2a) was performed in triplicate on three biological replicates. Staining of DF1 and DF3 with stress-granule markers in mES cells (Extended Data Fig. 2b, c) was performed in duplicate on biological replicates. Staining of DF2 and TIAR in HEK293, U2OS and NIH3T3 cells (Extended Data Fig. 2d–f) was performed in duplicate on biological replicates. Western blot of CRISPR–Cas9-edited NeonGreen–YTHDF2 edited cells (Extended Data Fig. 2g) was performed in duplicate on technical replicates. Imaging of NeonGreen–DF2 stress-induced granules (Extended Data Fig. 2h) was performed in duplicate on biological replicates. Fluorescence recovery after photobleaching in NeonGreen–DF2 cells (Fig. 2c) was performed on three granules in one biological sample. DF2 and EDC4 co-staining in mES cells (Fig. 1d) was performed in duplicate on biological replicates. DF2 staining after puromycin or actinomycin D treatment (Extended Data Fig. 2j) was performed in duplicate on biological replicates. TLC of stressed NIH3T3 cells (Fig. 4a, Extended Data Fig. 2k, l, Extended Data Fig. 4b) was performed in triplicate on three biological replicates for control conditions, and quadruplicate for four biological replicates under stress conditions. Staining of DF2 in stressed wild-type and *Mettl14*-knockout mES cells (Fig. 3a, b, Extended Data Fig. 4b) was performed in triplicate on three biological replicates. TLC of poly(A)-purified mRNA from wild-type and *Mettl14*-knockout mES cells (Extended Data Fig. 3a) was performed in duplicate on technical replicates. Transfection of NeonGreen–DF constructs (Fig. 3c) was performed in duplicate on biological replicates. Staining of EDC4 and DF2 in wild-type and *Mettl14*-knockout mES cells (Fig. 3d) was performed in duplicate on biological replicates. Validation of stress granule isolation from NIH3T3 cells (Extended Data Fig. 4) was performed in duplicate on biological replicates. Cumulative distribution plots of m^{6}A-mRNAs from stress granules in U2OS cells (Fig. 4b) were based on average values of $\log_2$-transformed RNA fold change generated from three biological replicates. Cumulative distribution plots of m^{6}A-mRNAs from isoxazole-induced neuronal RNA granules (Extended Data Fig. 4c) were based on average values of $\log_2$-transformed RNA fold change generated from three biological replicates. Cumulative distribution plots of m^{6}A-mRNAs from stress granules in NIH3T3 cells (Extended Data Fig. 4d) were based on average values of $\log_2$-transformed RNA fold change generated from three biological replicates. smFISH on mRNAs in mES cells (Fig. 4c, d) were performed in duplicate on biological replicates. Analysis of RNA-seq from wild-type mES cells (Fig. 4e, Extended Data Fig. 5a) were performed on average $\log_2$-transformed fold change values from four biological replicates. Total Ribo-seq coding sequence reads (Extended Data Fig. 5b) are from four biological replicates in each condition. Analysis of translational efficiency in wild-type mES cells compared with *Mettl14*-knockout cells before stress (Fig. 4f) was performed on four and three biological replicates, respectively. Analysis of translational efficiency in wild-type mES cells compared with *Mettl14*-knockout cells 1 h after stress (Fig. 4g) was performed on four and two biological replicates, respectively. Translational recovery experiments (Extended Data Fig. 5c) were performed in duplicate on biological replicates. Pearson's correlation coefficients for Ribo-seq reads (Extended Data Fig. 5d) were performed on the top two biological replicates, which were determined by samples with the highest percentage of mapped reads to the coding region.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The RNA sequencing (Fig. 4e, Extended Data Fig. 5a) and ribosome profiling (Fig. 4f, g, Extended Data Fig. 5b, d) data reported in this paper have been deposited in the NCBI Gene Expression Omnibus under accession number GSE125725. All other data are available from the corresponding author upon reasonable request.

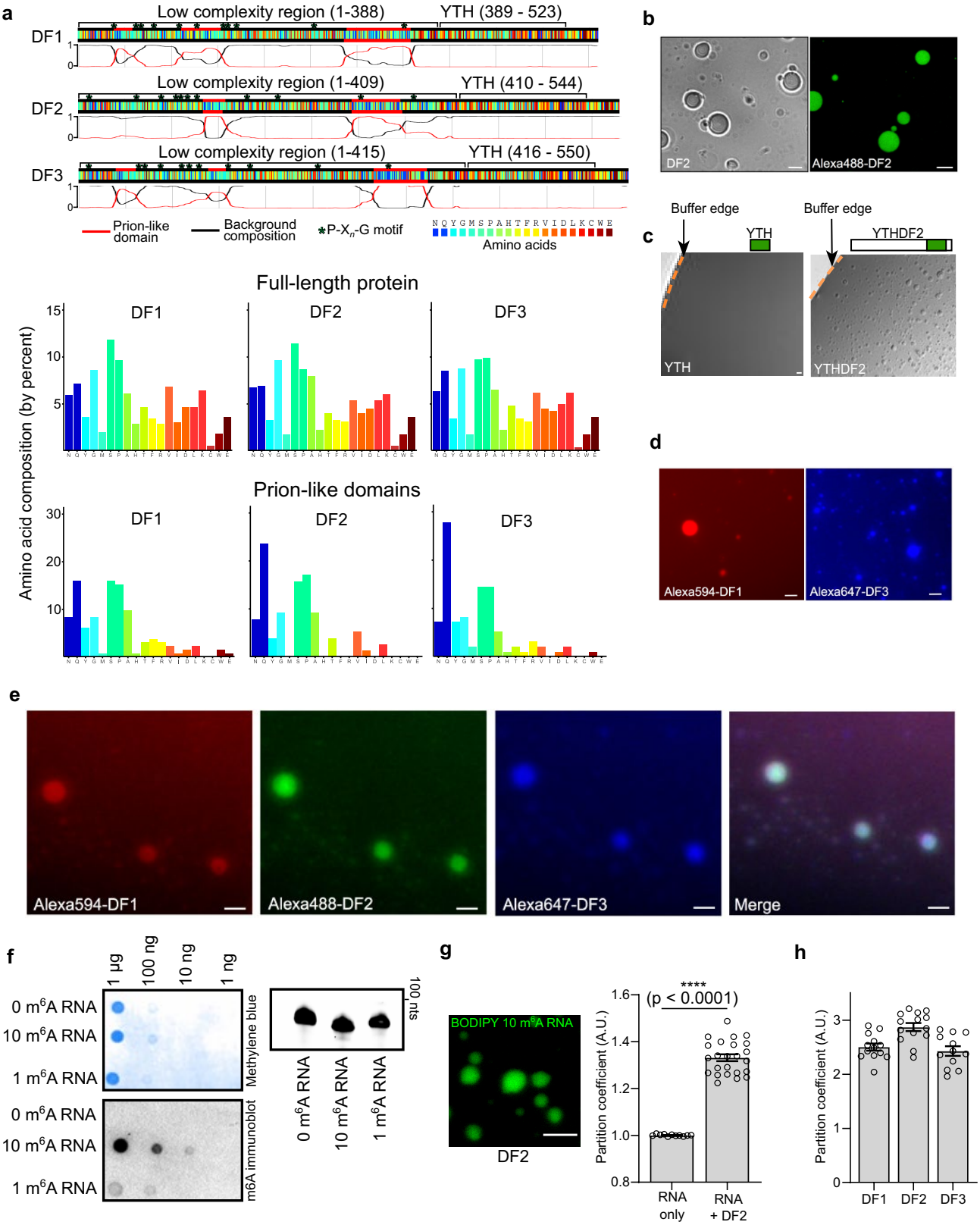
33. Knuckles, P. et al. Zc3h13/Flacc is required for adenosine methylation by bridging the mRNA-binding factor Rbm15/Spenito to the m^{6}A machinery component Wtap/Fl(2)d. *Genes Dev.* **32**, 415–429 (2018).
34. Wen, J. et al. Zc3h13 regulates nuclear RNA m^{6}A methylation and mouse embryonic stem cell self-renewal. *Mol. Cell* **69**, 1028–1038.e1026 (2018).
35. Patil, D. P. et al. m^{6}A RNA methylation promotes XIST-mediated transcriptional repression. *Nature* **537**, 369–373 (2016).
36. Xu, C. et al. Structural basis for the discriminative recognition of N^6-methyladenosine RNA by the human YT521-B homology domain family of proteins. *J. Biol. Chem.* **290**, 24902–24913 (2015).
37. Zhang, H. et al. RNA controls polyQ protein phase transitions. *Mol. Cell* **60**, 220–230 (2015).
38. Zhong, S. et al. MTA is an *Arabidopsis* messenger RNA adenosine methylase and interacts with a homolog of a sex-specific splicing factor. *Plant Cell* **20**, 1278–1288 (2008).
39. Hoover, D. M. & Lubkowski, J. DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* **30**, e43 (2002).
40. Li, H. et al. Design and specificity of long ssDNA donors for CRISPR-based knock-in. Preprint at https://www.biorxiv.org/content/10.1101/178905v1 (2017).
41. Linder, B. et al. Single-nucleotide-resolution mapping of m^{6}A and m^{6}Am throughout the transcriptome. *Nat. Methods* **12**, 767–772 (2015).
42. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
43. McGlincy, N. J. & Ingolia, N. T. Transcriptome-wide measurement of translation by ribosome profiling. *Methods* **126**, 112–129 (2017).
44. Xiao, Z., Zou, Q., Liu, Y. & Yang, X. Genome-wide assessment of differential translations with ribosome profiling data. *Nat. Commun.* **7**, 11194 (2016).
45. Mi, H. et al. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. *Nucleic Acids Res.* **38**, D204–D210 (2010).
46. Lancaster, A. K., Nutter-Upham, A., Lindquist, S. & King, O. D. PLAAC: a web and command-line application to identify proteins with prion-like amino acid composition. *Bioinformatics* **30**, 2501–2502 (2014).
47. Gilles, J. F., Dos Santos, M., Boudier, T., Bolte, S. & Heck, N. DiAna, an ImageJ tool for object-based 3D co-localization and distance analysis. *Methods* **115**, 55–64 (2017).
48. Hubstenberger, A. et al. P-body purification reveals the condensation of repressed mRNA regulons. *Mol. Cell* **68**, 144–157.e145 (2017).
49. Jain, S. et al. ATPase-modulated stress granules contain a diverse proteome and substructure. *Cell* **164**, 487–498 (2016).
50. Youn, J. Y. et al. High-density proximity mapping reveals the subcellular organization of mRNA-associated granules and bodies. *Mol. Cell* **69**, 517–532. e511 (2018).
51. Markmiller, S. et al. Context-dependent and disease-specific diversity in protein interactions within stress granules. *Cell* **172**, 590–604.e513 (2018).
52. Tsherniak, A. et al. Defining a cancer dependency map. *Cell* **170**, 564–576.e516 (2017).

Acknowledgements We thank members of the Jaffrey laboratory for comments and suggestions; members of the Epigenomics, Optical Microscopy and Imaging & Flow Cytometry Weill Cornell Cores for their assistance; and J. Hanna and S. Geula for generously providing *Mettl14*-knockout and control mES cell lines. This work was supported by NIH grants R01DA037755 (S.R.J.), F32CA22104-01 (B.F.P.), R01DK114131 (J.H.L.) and T32CA062948 (A.-O.G.) and an American-Italian Cancer Foundation fellowship (S.Z.).

Author contributions S.R.J., R.J.R., A.O.-G., S.Z. and P.K. designed the experiments. R.J.R. and P.K. carried out stress-granule-staining experiments; S.N., J.H.L. and H.K. prepared stress granules; R.J.R. and P.K. performed assays related to phase separation and stress-granule formation; P.K. performed puromycin-labelling assays, R.J.R. performed quantification of stress granules and smFISH; R.J.R., P.K. and S.Z. analysed ribosome-profiling data; A.O.-G. and R.J.R. performed analysis of stress-granule transcriptomes; B.F.P. performed m^{6}A measurements and D.P.P. made DF expression constructs. S.Z. performed CRISPR knock-in; R.J.R. and S.Z. performed in-cell FRAP experiments. R.J.R. and S.Z. prepared figures relating to ribosome-profiling data. R.J.R., S.Z. and P.K. prepared the remaining figures. S.R.J. wrote the manuscript with input from all authors.

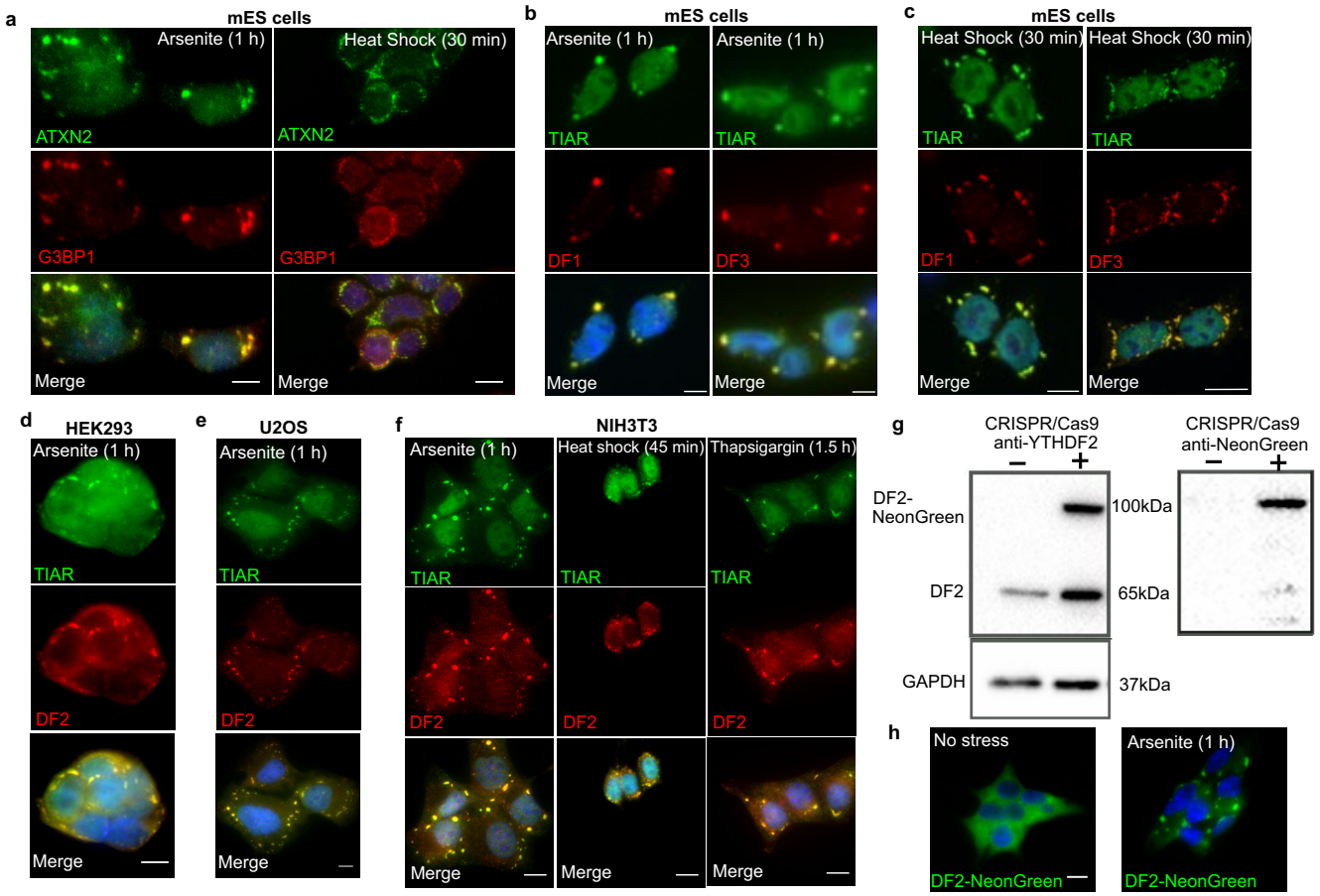
Competing interests S.R.J. is scientific founder of, advisor to, and owns equity in Gotham Therapeutics.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1374-1.
Correspondence and requests for materials should be addressed to S.R.J.
Peer review information *Nature* thanks Richard Kriwacki and Tanja Mittag for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.



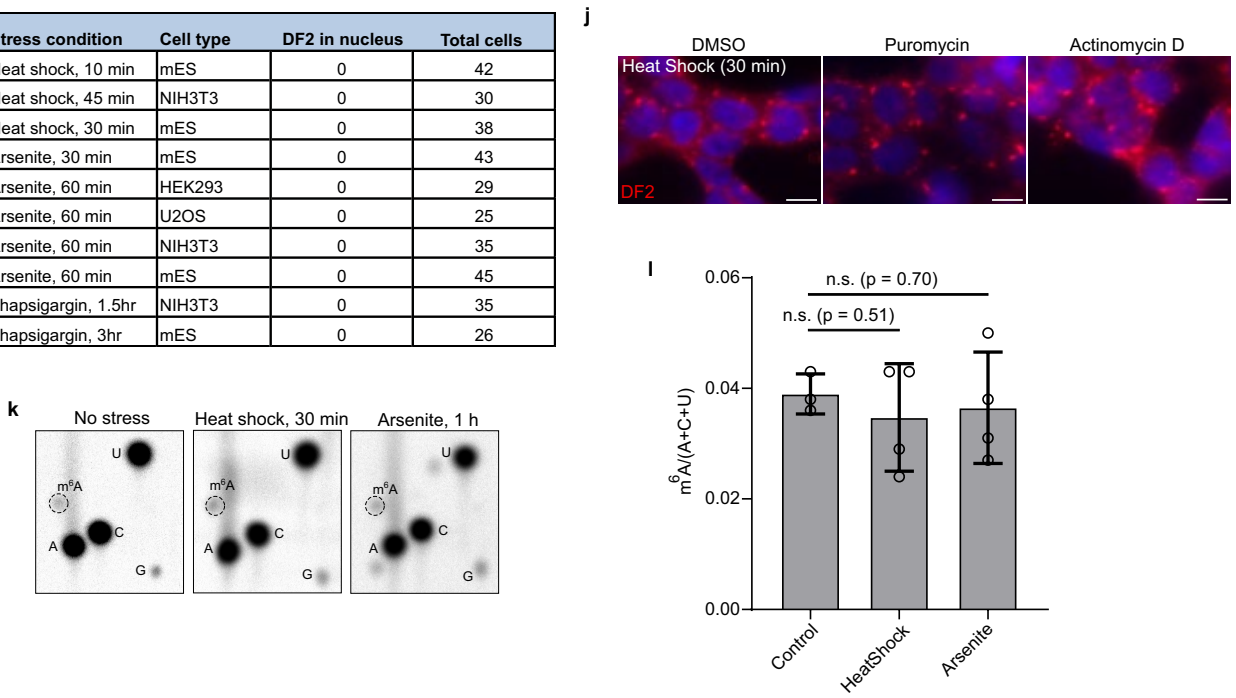
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Fluorescent labelling of DF2 does not affect the formation of liquid droplets. a, DF1, DF2 and DF3 exhibit high sequence homology. Shown is a colour-coded schematic representation of the aligned amino acid sequence and corresponding prion-like domain disorder propensity plots (red and black traces) for DF1, DF2 and DF3 generated using the PLAAC (prion-like amino acid composition) tool[46]. The *y* axis of the plot represents prion-like regions (1) and regions of background amino acid composition (0). The low-complexity domain is a region of approximately 40 kDa that contains glutamine-rich prion-like domains and an abundance of disorder-promoting residues such as proline, glycine, serine, alanine and asparagine. These domains are also enriched with multiple P-X_n-G motifs that are known to be associated with lower critical solution temperature[7]. The YTH domain (about 15 kDa) exhibits high sequence identity between the paralogues, and all YTH domains show identical binding to m^6A without preference for any specific sequence context surrounding m^6A[4]. The high degree of sequence identity suggests that these proteins might function redundantly in stress-granule formation and phase separation. Amino acid compositions of the full-length DF proteins and their prion-like domains are shown in the bar charts at the bottom of the panel. **b**, Liquid-droplet formation for Alexa488-labelled DF2. The goal of this experiment is to confirm that labelling DF2 with Alexa488 does not affect liquid-droplet formation. Indeed, before labelling DF2 with Alexa488, DF2 protein droplets were readily detectable by differential interference contrast microscopy (left). After labelling, Alexa488-labelled DF2 protein droplets are still observed by fluorescence microscopy (right). These data indicate that the labelling protocol does not impair droplet formation by DF2. Images are taken from different protein preparations. Experiments were performed in duplicate. Scale bar, 10 μm. **c**, The intrinsically disordered domain of DF2 is required for the phase separation of DF2. Bright-field microscopic images of recombinant DF2 lacking the N-terminal intrinsically ordered domain (YTH) and full length DF2 are shown (a schematic of the domain representation is shown above the image). The edge of the buffer (buffer/air interface) is shown with a dashed line. Whereas the full-length YHTDF2 (75 μM) can undergo phase separation, at the same concentration and in the same buffer conditions YTH cannot. This indicates that the intrinsically disordered domain is required for phase separation. Experiments were performed in duplicate. Scale bar, 10 μm. **d**, DF1 and DF3 undergo phase separation in vitro. Shown are fluorescence microscopy images of Alexa594–DF1 and Alexa647–DF3. DF1 and DF3 undergo phase separation in vitro as assessed by the formation of protein droplets. Experiments were performed in duplicate. Scale bar, 10 μm. **e**, DF1, DF2 and DF3 form protein droplets comprising all three proteins. Shown are fluorescence microscopy images of Alexa594–DF1, Alexa488–DF2 and Alexa647–DF3. Mixing the three recombinant proteins shows that these proteins can phase-separate together to form protein droplets that contain all three proteins. Experiments were performed in duplicate. Scale bar, 10 μm. **f**, Confirmation of in vitro-transcribed RNA abundance and methylation status. In vitro-transcribed RNAs were serially diluted (1:10) and stained for total RNA by methylene blue staining (top left) as well as m^6A abundance by immunoblotting using an anti-m^6A antibody (bottom left). RNA with no m^6A nucleotides gave no signal whereas RNAs with 10 m^6A nucleotides gave a significantly higher signal in the dot blot than those with 1 m^6A nucleotide. Additionally, in vitro-transcribed RNAs were analysed on a 15% denaturing gel, demonstrating the absence of degradation products (right). Experiments were performed in duplicate. **g**, Partition coefficients of fluorescently labelled m^6A RNAs with and without DF2. To determine the extent to which multi-m^6A-RNAs were recruited into DF droplets, we synthesized a 10-m^6A RNA with a 5′ BODIPY FL fluorescent tag and measured its partition coefficient in the presence of DF2 (7.5 μM, 20 mM HEPES pH 7.4, 300 mM KCl, 6 mM $MgCl_2$, 0.02% NP-40, 10% glycerol). Upon addition of 850 nM BODIPY–10-m^6A-RNA, fluorescent RNA-containing droplets appeared in minutes (left). A video of fluorescent DF2:BODIPY–10-m^6A-RNA coacervate droplet fusion is shown in Supplementary Video 2. Calculation of partition coefficients in comparison to background fluorescent-labelled RNAs[6] shows that m^6A mRNAs are enriched in DF2-containing droplets (right; RNA only, $n = 11$; RNA + DF2, $n = 24$, where n represents distinct droplets in biological replicates). The experiment was performed in duplicate. Bar heights represent mean partition coefficients and error bars represent s.e.m. $^{****}P < 0.0001$, two-sided Mann–Whitney test. Scale bar, 10 μm. **h**, The partition coefficient of DF proteins increases over time. In this experiment we measured the partition coefficient of DF1, DF2 and DF3 as shown in Fig. 1g. However, here we measured the values after 24 h, unlike the time point used in Fig. 1g (approximately 5 min). The partition coefficients are notably increased compared to the values measured in Fig. 1g. This suggests that droplet formation had not achieved equilibrium at the early time points used in Fig. 1g. Bar heights represent mean partition coefficients and error bars represent s.e.m. Experiments were performed in duplicate.



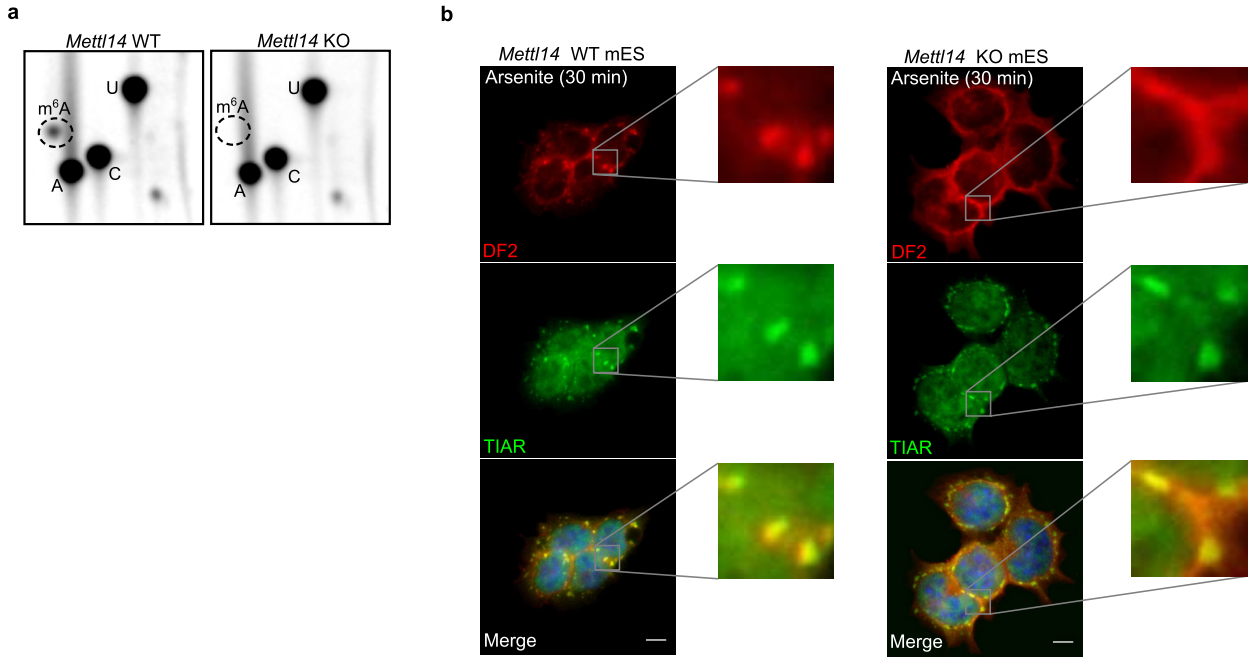
i

Stress condition	Cell type	DF2 in nucleus	Total cells
Heat shock, 10 min	mES	0	42
Heat shock, 45 min	NIH3T3	0	30
Heat shock, 30 min	mES	0	38
Arsenite, 30 min	mES	0	43
Arsenite, 60 min	HEK293	0	29
Arsenite, 60 min	U2OS	0	25
Arsenite, 60 min	NIH3T3	0	35
Arsenite, 60 min	mES	0	45
Thapsigargin, 1.5hr	NIH3T3	0	35
Thapsigargin, 3hr	mES	0	26

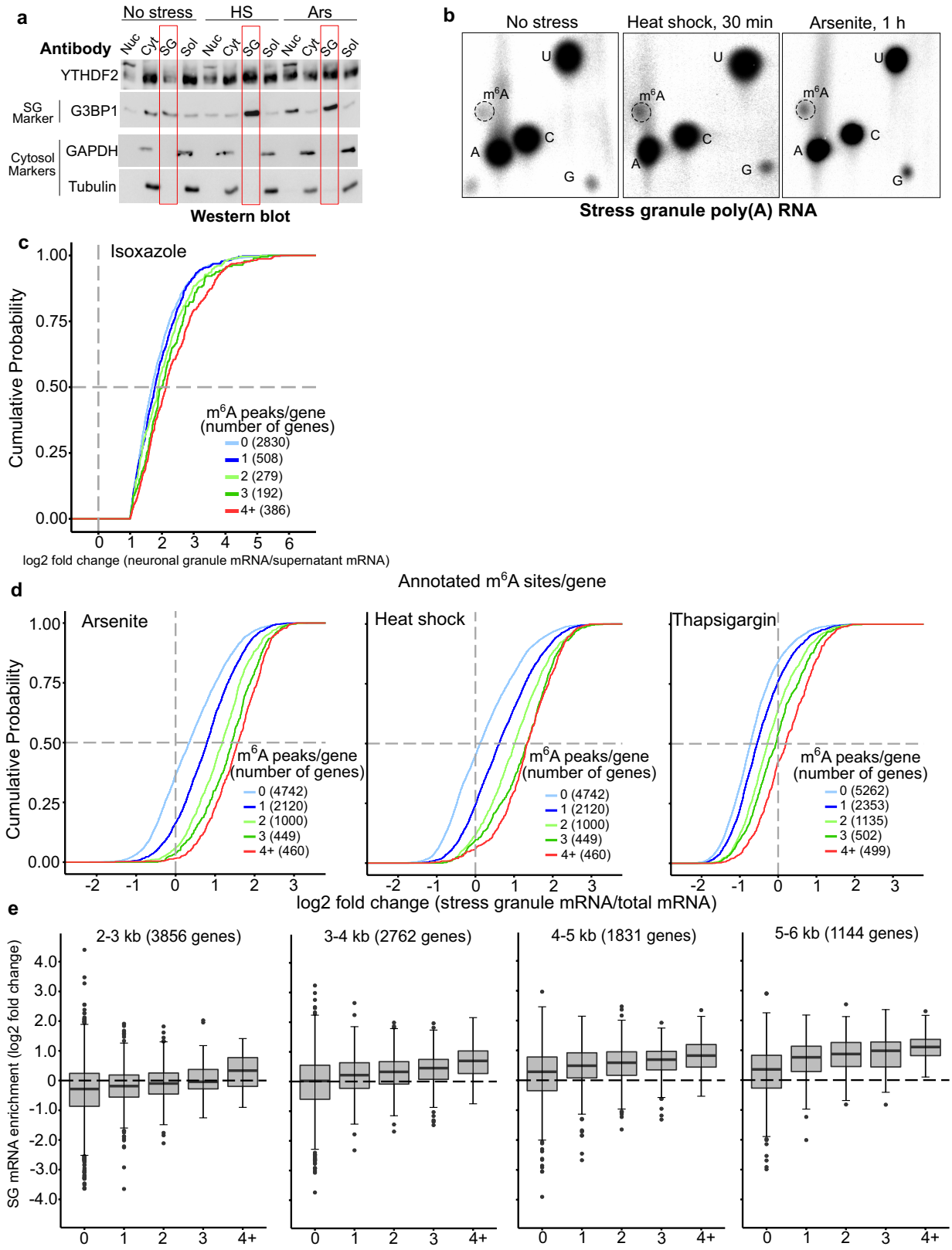


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | Assessing which stressors induce stress granules and the localization of DF2 proteins in diverse cell types. a, Oxidative stress and heat shock induce stress-granule formation in mouse ES cells. Stress-granule formation has not been extensively characterized in mouse ES cells. We therefore wanted to ensure that stress granule composition is the same in mouse ES cells compared to other cell types in which stress granules are more frequently studied. To test mouse ES stress granules, we stained with additional markers. Co-immunostaining with ATXN2 (green) and G3BP1 (red) after arsenite treatment (0.5 mM for 1 h) and heat shock (42 °C for 30 min) in mES cells showed clear labelling of stress granules. The overlay panel shows ATXN2 and G3BP1 overlap (yellow). Thus, stress granules in mouse ES cells appear to have similar markers as stress granules in other cell types. The experiment was performed in triplicate. Scale bar, 10 μm. **b**, **c**, DF1 and DF3 proteins relocalize to stress granules after heat shock and oxidative stress. DF1, DF2 and DF3 have high sequence similarity and show similar phase-separation properties. We therefore wanted to determine whether all these proteins associate with stress granules. Co-immunostaining was performed in mES cells with DF1 (red) or DF3 (red) with TIAR (green) after arsenite treatment (0.5 mM for 1 h) or heat shock (42 °C for 30 min). Along with DF2 shown in Fig. 2, DF1 and DF3 relocalize to stress granules treatment as visualized by the co-localization with TIAR. Scale bar, 10 μm. These findings are consistent with previous proteomic datasets of stress granules. A P-body proteome dataset[48] showed that DF2 was enriched in P-bodies. DF2 ranked 152 among 1,900 P-body-associated proteins by abundance. All DF proteins were identified in a group of around 300 stress-granule-enriched proteins in a proteomics study of stress granules[49]. In another study, in vivo proximity-dependent biotinylation (BioID)-labelling study of G3BP1 and other stress-granule markers showed interactions with all DF proteins[50]. Another APEX labelling study[51] of G3BP1 showed that the YTHDFs are 3 of the top 42 G3BP1-interacting proteins in the stress-granule proteome. Overall, these studies suggest that DF proteins are commonly seen in stress granules, and may be highly abundant relative to other stress-granule components. The experiment was performed in triplicate. **d**–**f**, DF2 relocalizes to stress granules after arsenite treatment in numerous cell types. The focus of this experiment was to determine whether DF relocalization to stress granules is likely to be a universal feature of stress granules. We therefore tested DF localization to stress granules in multiple cell types. Shown is co-immunostaining of HEK293 cells (**d**), U2OS cells (**e**), and NIH3T3 cells (**f**) with DF2 (red) and TIAR (green) after arsenite treatment (0.5 mM for 1 h) and heat shock (42 °C for 30 min). The overlay panel shows DF2 in stress granules based on its overlap with TIAR (yellow). The experiment was performed in duplicate. Scale bar, 10 μm. **g**, Confirmation of CRISPR–Cas9 knock in of NeonGreen–DF2. A western blot of HEK293T shows endogenous expression of NeonGreen–DF2. Note, only one allele contains the knock-in construct, accounting for the presence of unmodified DF2 in cells. **h**, Arsenite stress induces the localization of NeonGreen–DF2 into stress granules. We wanted to determine whether the ability of DF2 to undergo phase separation in vitro could be actively observed in cells. Unstressed HEK293T cells expressing NeonGreen-tagged DF2 protein show a diffuse cytoplasmic fluorescent signal. Upon arsenite stress (0.5 mM, 1 h), NeonGreen–DF2 phase-separates into stress granules. This confirms the ability of NeonGreen–DF2 to undergo phase separation in cells in response to stress. The experiment was performed in triplicate. Scale bar, 10 μm. **i**, Relocalization of DF2 to the nucleus does not occur after various stresses in various cell types. Because DF2 has been reported to relocalize to the nucleus 2 h after heat shock[13], we wanted to determine whether any nuclear relocalization occurs in our experiments, which were performed immediately after stress. The 'Stress condition' column indicates the type and length of stress applied. The 'Cell type' column indicates the type of cell that was stressed. The 'DF2 in nucleus' column denotes the number of cells that were found to have DF2 in the nucleus immediately after stress. The 'Total cells' column indicates the number of cells that were examined for DF2 nuclear relocalization in each experimental condition. In all conditions, there was no cell that showed nuclear DF2 localization. Thus, DF2 localization is primarily in cytosolic stress granules at the time at which the stress is terminated. DF2 was not observed to relocalize to the nucleus at any time point or after any stress, including the 2-h post-heat-shock conditions described previously[13]. **j**, DF2 relocalization to stress granules does not require new mRNA or protein synthesis. We wanted to know whether an increase in DF2 expression or new m^{6}A formation could be required for the formation of stress granules after heat shock. To test this, we blocked protein synthesis with puromycin and blocked new transcription with actinomycin D. Actinomycin D blocks m^{6}A formation because m^{6}A formation occurs co-transcriptionally[17,18]. The fluorescence micrographs show DF2 immunostaining in HEK293T cells treated with DMSO (left), puromycin (10 μg ml^{-1}, middle), and actinomycin D (2.5 μg ml^{-1}, right) for 15 min before and during incubation at 42 °C for 30 min. The ability of DF2 to relocalize to stress granules when transcription (actinomycin D) and translation (puromycin) was arrested was assessed by immunofluorescence staining for DF2. In each case, the formation of stress granules was unaffected, indicating that no new protein synthesis or new methylation is required for stress-granule formation. The time course of stress-granule formation is rapid, making it unlikely that new protein synthesis or methylation is involved in stress-granule formation. Additionally, heat shock is normally associated with inhibited transcription and translation, which further suggests that new protein synthesis and RNA methylation is unlikely to occur in the time course of stress-granule formation. On the basis of all this data, we propose that stress granule formation probably utilizes pre-existing patterns of m^{6}A seen in unstressed cells to mediate the formation of stress granules. The experiment was performed in duplicate. Scale bar, 10 μm. **k**, **l**, m^{6}A levels are not significantly altered immediately after arsenite and heat-shock stress in NIH3T3 cells. We wanted to test whether m^{6}A levels in mRNA transcripts were altered as a result of cellular stress. NIH3T3 cells were subjected to arsenite (0.5 mM, 1 h) or heat-shock stress (43 °C, 45 min) and total RNA was extracted immediately after stress treatment. Total RNA was further purified by poly(A) selection to specifically assay m^{6}A levels in mRNA transcripts. TLC[38] revealed that there was no significant increase in m^{6}A levels within poly(A) mRNA immediately after either stress condition in three biological replicates (see **l**). This indicates that cellular stress does not induce an increase or decrease in m^{6}A over the time frame examined. Experiments were performed in duplicate. Bar heights in **l** represent mean and error bars represent s.e.m. Three biological replicates ($n = 3$) were analysed in the control, and four biological replicates ($n = 4$) were analysed after heat-shock and arsenite stress. Stress m^{6}A/(A+C+U) ratios were analysed with a two-sided Student's *t*-test.

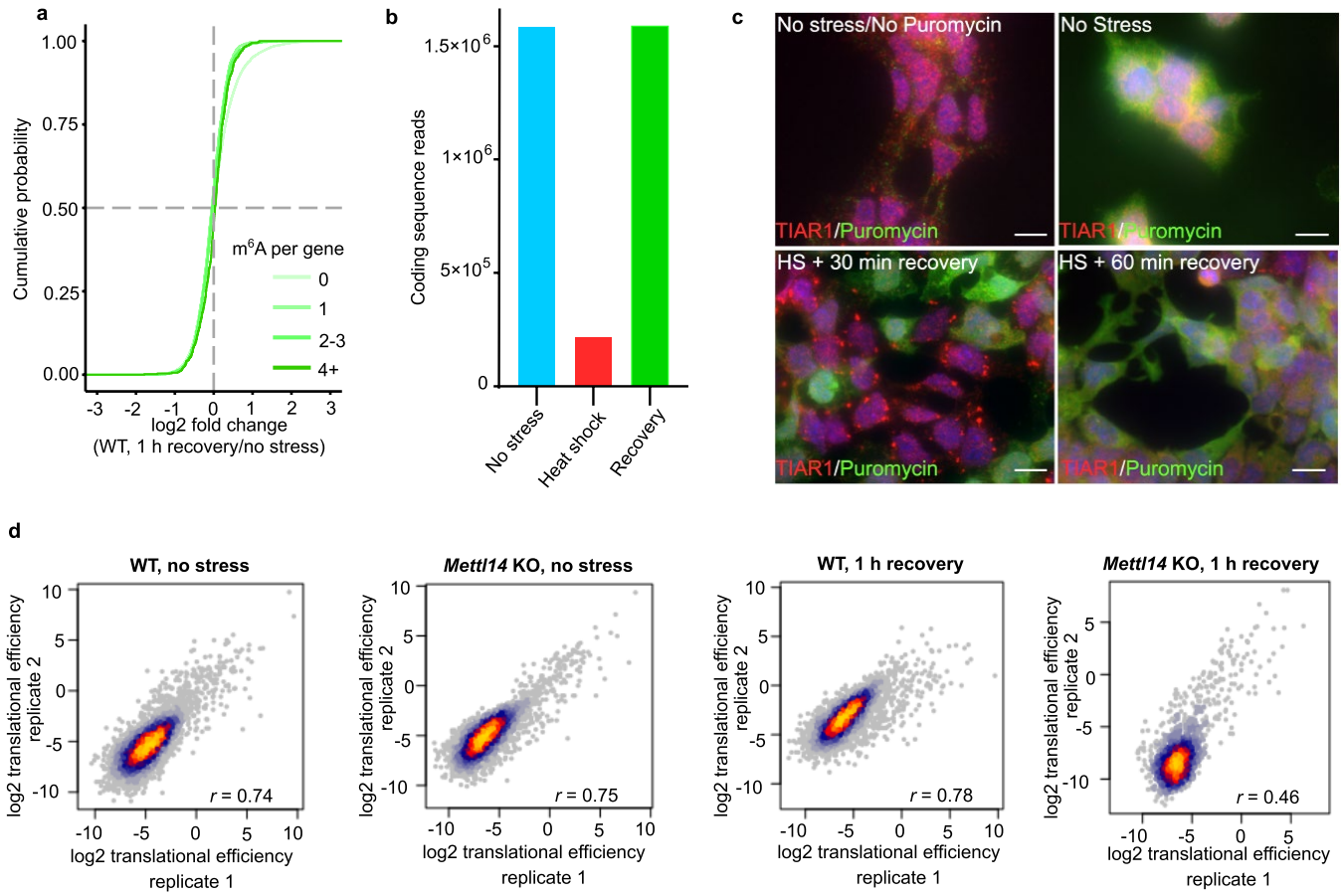


Extended Data Fig. 3 | Confirmation of the *Mettl14*-knockout model and DF2 phase-separation into P-bodies in mES cells. a, *Mettl14*-knockout (KO) mES cells are depleted in m^{6}A RNA. We sought to independently confirm the depletion of m^{6}A from mRNA in these cells, which were previously shown to have a 99% reduction in m^{6}A[32]. The TLC assay selectively quantifies m^{6}A in a G–A–C context, thereby reducing the possibility of contamination of m^{6}A from ribosomal RNA or small nuclear RNA, which are quantified in an A–A–C or a C–A–G context, respectively[38]. The protocol was performed as described previously[38]. Indicated in the TLC chromatograms are the relative positions of m^{6}A (dotted circle) and adenosine (A), cytosine (C) and uracil (U). The left and right panels show radiochromatograms obtained from 2D-TLC of poly(A) RNA from wild-type and *Mettl14*-knockout cells. No m^{6}A is detectable in the poly(A) RNA derived from *Mettl14*-knockout cells, which confirms the efficiency of m^{6}A depletion in these cells. Experiments were performed in duplicate. mES cells are used here because m^{6}A depletion can be readily achieved in *Mettl14*-knockout mES cells without impairing viability[32]. By contrast, m^{6}A depletion cannot be readily achieved in immortalized cell lines as both *Mettl3* and *Mettl14* are essential for nearly all cell lines[52]. **b**, DF2 partitioning into stress granules induced by arsenite is delayed in m^{6}A-deficient cells. This delay is similar to that shown in stress granules induced by heat shock as seen in Fig. 3a. The experiment was performed in triplicate.

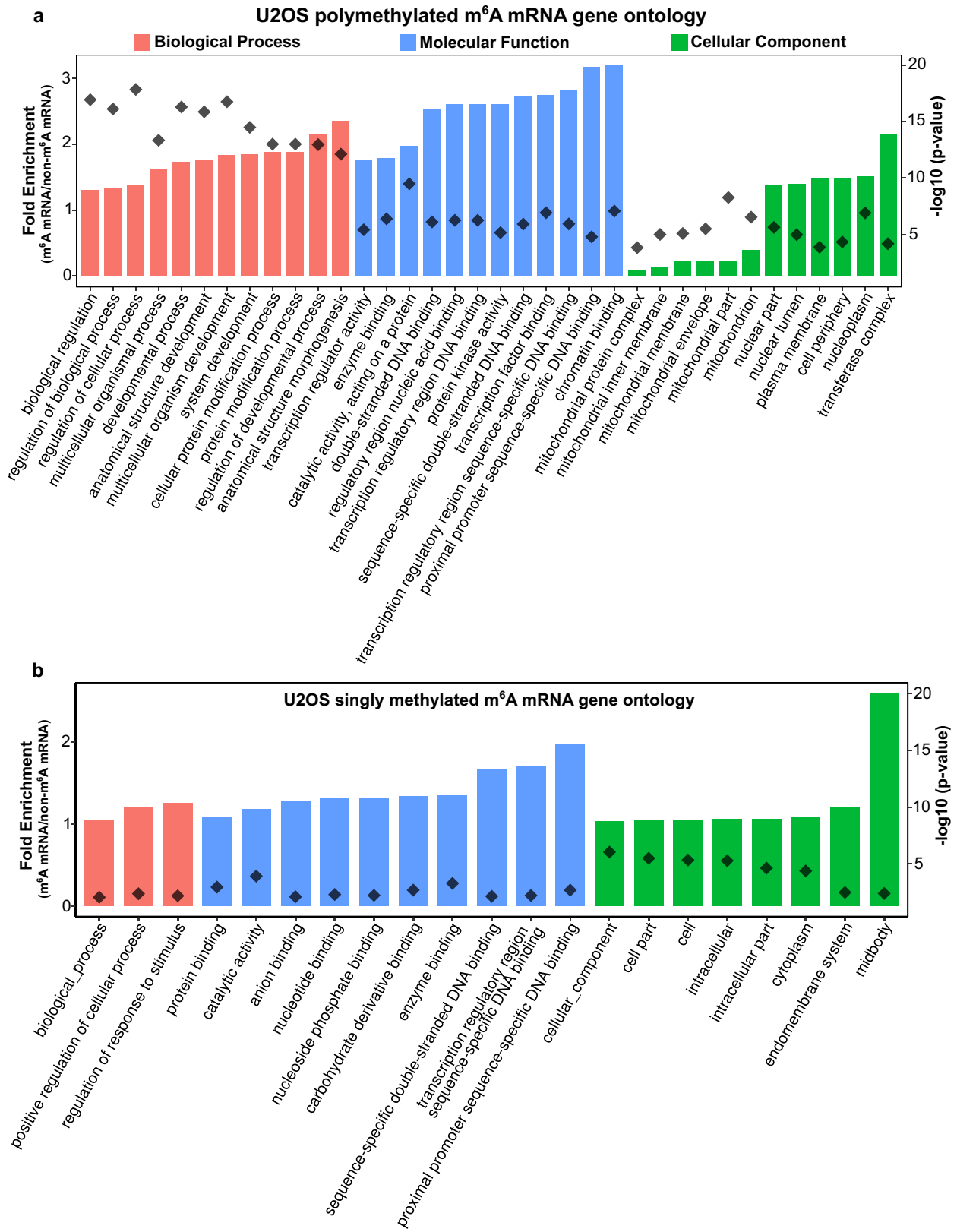


Extended Data Fig. 4 | See next page for caption.

Extended Data Fig. 4 | The number of m^{6}A nucleotides is correlated with stress-granule enrichment independent of transcript length. **a**, DF2 is enriched in stress granules after stress. Nuclear (Nuc), cytosolic (Cyt), insoluble RNA-granule enriched (SG, red boxed lanes), and soluble (Sol) protein fractions were isolated from stressed NIH3T3 cells as described previously[24]. G3BP1 was used as a stress-granule marker. GAPDH and tubulin were used as cytosolic and soluble fraction markers. Under non-stressed conditions, DF2 is most abundant in the cytoplasmic and soluble protein fractions. However, upon both arsenite and heat-shock stress, the highest levels of DF2 are found in the RNA granule (insoluble) fraction, which indicates that diverse stresses cause the partitioning of DF2 from the cytosol into stress granules. The experiment was performed in duplicate. **b**, m^{6}A levels are increased in the mRNAs in the insoluble stress-granule-enriched fraction after cellular stress in NIH3T3 cells. Shown are representative TLC plates analysing m^{6}A levels in mRNAs in the stress-granule fraction from the analysis presented in Fig. 4a. Representative plates from the cytosolic fraction are shown in Extended Data Fig. 2l. Experiments were performed in duplicate. **c**, m^{6}A number correlates with mRNA enrichment in RNA granules in mouse neurons. In these experiments, we used mRNA enrichment data (RNA granule versus supernatant) derived from a study of isoxazole-induced RNA granules in mouse brain[22]. Enriched mRNAs are defined by a $\log_2$-transformed fold change of more than 1 in the RNA granule fraction relative to the supernatant fraction. A cumulative distribution plot of mRNA enrichment was performed for mRNAs classified by the number of called m^{6}A peaks per gene based on single-nucleotide resolution m^{6}A maps generated in the mouse brain. Transcripts that contain multiple m^{6}A peaks are enriched in RNA granules relative to non-methylated or singly methylated mRNAs. The original experiments were performed in triplicate. **d**, The number of m^{6}A sites in an mRNA correlates with its enrichment in stress granules in NIH3T3 cells. In these experiments, we used a dataset of relative mRNA enrichment data (stress granule versus cytoplasm) generated in a previous study[24]. Assignment of the number of m^{6}A sites in each transcript was based on a mouse embryonic fibroblast MeRIP-seq dataset obtained previously[32]. Analysis was performed as in Fig. 4c. Polymethylated mRNAs show greater enrichment in stress granules than non-methylated or singly methylated mRNAs for each stress condition. Experiments were performed in triplicate. **e**, Examination of the effect of m^{6}A on mRNA enrichment using controlled transcript size. Because transcript length positively correlates with stress-granule mRNA enrichment (see refs [23,24]), we wanted to control for this feature in our analysis. The same m^{6}A maps and RNA-seq data from U2OS cells that were used to generate Fig. 4c were used here. Transcripts were binned on the basis of their annotated transcript length (2–3 kb, 3–4 kb, 4–5 kb, 5–6 kb) and further sorted on the basis of the number of annotated m^{6}A sites in each transcript. We found that mRNAs annotated with fixed lengths each showed increased enrichment based on the number of mapped m^{6}A sites. The number of m^{6}A per transcript was a positive predictor of transcript enrichment in stress granules even when controlling for transcript length. The centre of the box plot represents the median $\log_2$-transformed fold change, the boundaries contain genes within a quartile of the median, the whiskers represent genes in the upper and lower quartiles, and the outliers are presented as dots.

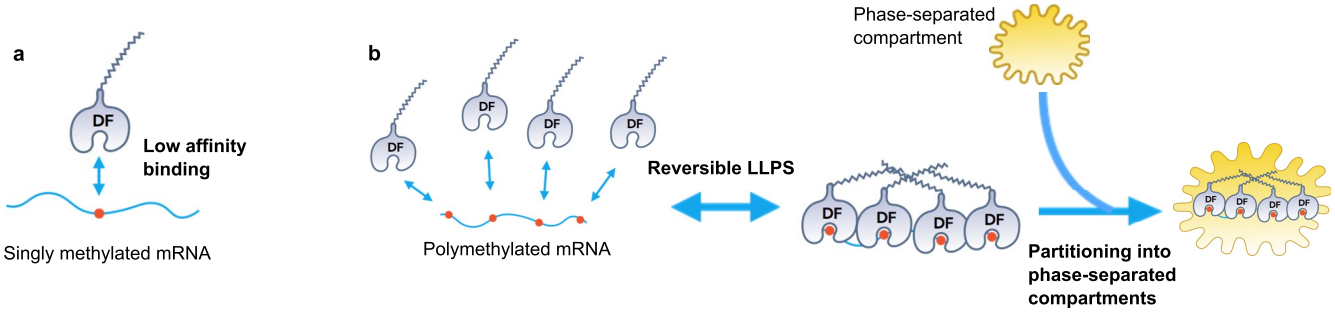


Extended Data Fig. 5 | Detection of translation in mES cells after stress. **a**, m^{6}A-mRNA transcript abundance is similar before stress and after stress. We wanted to understand whether mRNA transcript abundance was altered as a result of DF mobilization in mES cells after heat shock. In Fig. 4a, we examined RNA expression before heat shock and compared it to mRNA levels after 30 min of heat shock. Here, we allowed the cells to recover for 1 h, reasoning that this additional time might allow for DF-mediated mRNA degradation. As in Fig. 4a, we performed RNA-seq on wild-type mES cells before heat shock and after stress, measured after cells were returned to 37 °C for 1 h. The same m^{6}A annotation strategy was used as in Fig. 4a. As can be seen, the levels of m^{6}A in an mRNA is not correlated with an alteration in mRNA abundance. The $\log_2$-transformed fold change values represent the average of four biological replicates. **b**, Raw counts for ribosome protected fragments. Ribosome-protected fragments were collected from mES cells before stress, immediately after heat shock (42 °C, 30 min) and 1 h after heat shock. The number of ribosome-protected fragments isolated from cells immediately after heat shock was substantially lower than the number of ribosome-protected fragments isolated before stress and 1 h after stress. This indicates that translation is globally suppressed during the heat shock. As a result of the few ribosome-protected fragments during heat shock, translational efficiency could not be calculated during heat shock. Bar heights represent the totals from four biological replicates in each condition. **c**, Translation recovers 1 h after heat shock in mES cells. Here we assessed the amount of time needed for translation to be detected after heat shock. mES cells were subject to heat shock for 30 min at 42 °C and translation was assessed at different time points after the cessation of heat shock. Translation was monitored by labelling nascent peptides with puromycin. Puromycin was added to cells for 10 min. Immunostaining with an antibody against TIAR and puromycin provides a correlation between the presence of stress granules and the translation state. Non-stressed cells that were not treated with puromycin are shown as a control to establish the background signal (upper left). Unstressed cells treated with puromycin show robust translation (green). At a recovery time of 30 min, most cells still contain stress granules (TIAR, red) and translation is absent except in the few cells that lack stress granules. However, at 1 h, translating cells can be readily detected on the basis of puromycin immunoreactivity reactivity. Less than 50% of cells exhibit stress granules. On the basis of these experiments, we used 1 h as the time point for our ribosome-profiling experiments. The experiments were performed in duplicate. Scale bar, 10 μm. **d**, Comparison of the two biological replicates with the highest percentage of CDS-mapped reads in each condition for ribosome-profiling experiments. Shown are Pearson's correlation plots for the replicates used in the translational efficiency analysis shown in Fig. 4f, g.



Extended Data Fig. 6 | See next page for caption.

Extended Data Fig. 6 | Gene ontology of polymethylated and singly methylated mRNAs in U2OS cells. a, **b**, Gene ontology of m^{6}A-mRNAs in U2OS cells. U2OS RNA-seq data from ref. [23] and Me-RIP-seq data from ref. [31] were used in a gene ontology analysis for polymethylated m^{6}A-mRNAs in U2OS cells. Polymethylated mRNAs were defined as all mRNAs that had four or more annotated m^{6}A sites in the MeRIP-seq dataset that were identified in the U2OS RNA-seq ($n = 652$). Singly methylated mRNAs were defined as mRNAs that had one annotated m^{6}A site, with the same criteria ($n = 2,896$). Gene ontology was performed using the PANTHER gene ontology database. The biological process gene ontology search showed enrichment of regulatory and developmental-associated genes in the polymethylated group. The molecular function gene ontology search showed enrichment of protein, ion, enzyme and adenylyl-binding proteins, and de-enrichment of ribosome structural components. The cellular component gene ontology search showed de-enrichment of mitochondrial and ribosomal proteins, and enrichment of components of the nucleus and cell membrane. mRNAs that met the same inclusion criteria but had no annotated m^{6}A sites were used as the reference category for the gene ontology analysis ($n = 5,956$). Fold enrichment scores for each gene ontology category are indicated by the coloured bars and correspond to the left *y* axis. *P* values for each gene ontology category are indicated by the dark grey diamonds and correspond to the right *y* axis. *P* values were determined with Fisher's exact test and a Bonferroni correction was performed for multiple hypothesis testing.



Extended Data Fig. 7 | Model of how the properties of m^{6}A-containing mRNAs are determined by their phase separation into intracellular phase-separated compartments. a, Depicted is the binding of DF proteins to singly methylated mRNAs. DF proteins show low-affinity interactions with m^{6}A-containing mRNAs. Affinities typically range between 0.9–1.1 μM for DF1, DF2 and DF3[4,36]. These low affinities suggest that DF proteins would not be able to form a stable bimolecular interaction with singly methylated RNA. The low affinity can now be understood in the context of phase separation. Their weak interactions with RNA are probably stabilized by interactions between their low-complexity domains, and subsequent phase separation. Notably, all m^{6}A sites in cytosolic mRNAs appear to have an equal propensity to bind each DF protein[35]. Thus, any m^{6}A residue may be sufficient to enhance the phase-separation potential of an mRNA. However, higher-level information—such as the spacing of m^{6}A sites—as well as other mRNA-bound proteins with low-complexity domains, may further affect the efficiency of phase separation. **b**, Polymethylated mRNAs bind multiple DF proteins leading to phase separation. When multiple DF proteins bind to a polymethylated mRNA, their interactions with the mRNA are stabilized by DF–DF interactions mediated by their low-complexity domains. These complexes may be reversible and undergo an assembly–disassembly equilibrium. However, if P-bodies, neuronal granules or stress granules are present, the DF–mRNA complexes can partition into these structures. m^{6}A-mRNA is then regulated by the regulatory proteins and functional properties of these distinct structures. If an mRNA has a single m^{6}A site, the mRNA can still partition into phase-separated structures, especially if RNA–RNA interactions or other RNA–protein interactions can facilitate phase separation. The DF low-complexity domain could interact with these non-DF proteins to enhance mRNA partitioning. Overall, interactions between DF proteins and m^{6}A-mRNAs probably lower the saturation concentration for their incorporation within stress granules, enhancing their partitioning over non-methylated mRNAs.

nature research

Corresponding author(s): DBPR 2019-01-00815A

Last updated by author(s): May 30, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to generate data in this article.
Data analysis	For NGS analysis: UNIX-based (Flexbar v2.5, pyCRAC v1.3.2, STAR aligner 2.6.0); R packages (xtail 1.1.5, DESeq2_1.18.1) For image analysis: Fiji (ImageJ) 1.52n, GraphPad Prism 8 v8.01, Nikon NIS-Elements Viewer v3.22.15 For GO analysis: PANTHER Classification System v14.0 (www.pantherdb.org)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated in this study will be available at GSE125725 upon article release. These data were analyzed in Figure 4e-g and Extended Data 5a, b, d.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No sample-size calculations were performed. For all biological data where representative data are depicted, experiments were performed in at least two biological replicates. In cases where biological replicates were not in close agreement, technical replicates of each condition were re-analyzed. In cases where variability in biological duplicates following technical replicate analysis was substantial, an additional biological replicate was analyzed with a corresponding technical replicate. Regardless of outcome, the conclusions that could be based on this triplicate analysis were accepted. NGS experiments with mES cells used four biological replicates. This was done due to the low number of ribosome protected fragments and that were detected in mES cells following heat shock.
Data exclusions	For the analysis of isoxazole-induced RNA granules in Extended Data 4c, genes with fewer than 1 log2 fold positive change in stress granule RNA enrichment are not included. This approach was used in the original analysis by Han, et. al. 2012, from which the data for our analysis was derived. For Ribo-seq data, low quality datasets, which we defined as runs which mapped to less than 50% CDS reads, were excluded from the final analysis of translational efficiency. This determination was made after the original analysis of Ribo-seq data. Inspection of metagene boundaries between the 5'UTR and CDS or CDS and 3'UTR showed that replicates with fewer than 50% CDS reads lacked a defined boundary for reads aligning to the CDS and/or UTRs and also lacked evidence of ribosome periodicity, a hallmark of high-quality libraries composed of ribosome-protected fragments. As such, we could not firmly conclude that the mapping in these replicates was of good quality, and we surmised that including these low-quality replicates may cause an incorrect interpretation of the data. We determined removal of these replicates appropriate since at least two replicates in each condition met the cutoff for inclusion and allowed for an interpretation of higher-quality data.
Replication	All experiments described were replicated at least once and no significant problems were encountered replicating any reported results.
Randomization	Samples were randomized for thin-layer chromatography experiments since experimental variability, even among technical replicates, can be substantial. All groups were allocated based on established biological differences that were either previously and/or independently verified. To our knowledge, no other experiments were performed in which randomization could have affected the interpretation of results.
Blinding	Investigators were blinded for thin-layer chromatography experiments and data analysis. Blinding was not possible for studies of WT and Mettl14 KO mES cells due to differences in cellular morphology that are clearly discernible by eye in fixed and living cells. To our knowledge, blinding in other experiments would be unlikely to affect the experimental outcome or interpretation of results.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	The following antibodies were used for immunofluorescence experiments: rabbit anti-TIAR (5137S, Cell Signaling Technology, Lot #1, 1:100), mouse anti-TIAR (Clone 6) (610352, BD Biosciences, Lots #5357680; 7219778, 1:100), mouse anti-Edc4 (H-12) (sc-376382, Santa Cruz Biotechnology, Lot #I0216, 1:100), mouse anti-Puromycin clone 12D10 (MABE343, Millipore Sigma, Lot

#2861354, 1:100), rabbit anti-G3BP1 (13057-2-AP, Proteintech, Lot #00047654, 1:100), mouse anti-ATXN2 (Clone 22) (611378, BD Biosciences, Lot #7341666, 1:100), rabbit anti-YTHDF1 (17479-AP, Proteintech, Lot #00044547, 1:100), rabbit anti-YTHDF2 (24744-1-AP, Proteintech, Lot #00053880, 1:100), rabbit anti-YTHDF3 (ab103328, Abcam, Lot #GR35115-39, 1:100), rabbit anti-IgG Alexa Fluor 594 (A11012, Invitrogen, Lot #1933366, 1:1000), mouse anti-IgG Alexa Fluor 488 (A11001, Invitrogen, Lot #1939600, 1:1000).

The following antibodies were used for immunoblotting experiments: rabbit anti-YTHDF2 (ARP67917_P050, Aviva Systems Biology, Lot #QC38405-43182, 1:1000), mouse anti-GAPDH (SC-365062, Santa Cruz Biotechnology, Lot #A2816, 1:5000), rabbit anti-IgG HRP (NA934V, GE Healthcare, Lot #16677077, 1:10000), mouse anti-IgG HRP (NA931V, GE Healthcare, Lot #16814909, 1:10000).

Validation

Antibodies used for IF:
rabbit anti-TIAR (5137S, Cell Signaling Technology) - manufacturer has validated use for IF in mouse and human
mouse anti-TIAR (Clone 6) (610352, BD Biosciences) - manufacturer has validated use for IF in mouse and human
mouse anti-Edc4 (H-12) (sc-376382, Santa Cruz Biotechnology) - manufacturer has validated use for IF in mouse and human
mouse anti-Puromycin clone 12D10 (MABE343, Millipore Sigma) - manufacturer has validated use for IF in mouse and human
rabbit anti-G3BP1 (13057-2-AP, Proteintech) - manufacturer has validated use for IF in mouse and human
mouse anti-ATXN2 (Clone 22) (611378, BD Biosciences) - manufacturer has validated use for IF in human. ATXN2 is a well-characterized stress granule marker, and this antibody has been used in IF studies previously in monkey (Nonis et al., 2008) and rat (Nonhoff et al., 2007). We found ATXN2 co-localized well with G3BP1 (a known stress granule marker) in both human cells and mouse embryonic stem cells (Extended Data 2a). ATXN2 localization to SGs is similar to what has been observed in previous stress granule studies (Kaehler et al. 2012) in human cells.
rabbit anti-YTHDF1 (17479-AP, Proteintech) - manufacturer has validated use for IF in mouse and human
rabbit anti-YTHDF2 (24744-1-AP, Proteintech) - manufacturer has validated use for IF in mouse and human
rabbit anti-YTHDF3 (ab103328, Abcam) - manufacturer has validated use for IHC in human. A previous publication validated the use of this antibody for IF in mouse (Yu et al., 2017 - Dynamic m6A modification regulates local translation of mRNA in axons)

Antibodies used for WB:
rabbit anti-YTHDF2 (ARP67917_P050, Aviva Systems Biology) - manufacturer has validated use for WB in human
mouse anti-GAPDH (SC-365062, Santa Cruz Biotechnology) - manufacturer has validated use for WB in human

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Mettl14 knockout mouse embryonic stem cells were a gift from J. Hanna and S. Geula (Weizmann Institute of Science). HEK293 (ATCC CRL-11268), U2OS (ATCC HTB-96), and NIH3T3 (ATCC CRL- 1658) were obtained from ATCC.
Authentication	Mettl14 knockout mES cells were previously described in Geula, et. al. 2015. HEK293, U2OS, and NIH3T3 cells were not authenticated.
Mycoplasma contamination	Mycoplasma staining was routinely tested with Hoechst staining. No mycoplasma were detected at any point in any cell line used in this study.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

LETTER

https://doi.org/10.1038/s41586-019-1372-3

Structure and assembly of the mitochondrial membrane remodelling GTPase Mgm1

Katja Faelber[1,10]*, Lea Dietrich[2,10], Jeffrey K. Noel[1], Florian Wollweber[3], Anna-Katharina Pfitzner[4], Alexander Mühleip[2], Ricardo Sánchez[5], Misha Kudryashev[5], Nicolas Chiaruttini[4], Hauke Lilie[6], Jeanette Schlegel[1], Eva Rosenbaum[1], Manuel Hessenberger[1], Claudia Matthaeus[1], Séverine Kunz[7], Alexander von der Malsburg[3], Frank Noé[8], Aurélien Roux[4], Martin van der Laan[3], Werner Kühlbrandt[2]* & Oliver Daumke[1,9]*

Balanced fusion and fission are key for the proper function and physiology of mitochondria[1,2]. Remodelling of the mitochondrial inner membrane is mediated by the dynamin-like protein mitochondrial genome maintenance 1 (Mgm1) in fungi or the related protein optic atrophy 1 (OPA1) in animals[3–5]. Mgm1 is required for the preservation of mitochondrial DNA in yeast[6], whereas mutations in the *OPA1* gene in humans are a common cause of autosomal dominant optic atrophy—a genetic disorder that affects the optic nerve[7,8]. Mgm1 and OPA1 are present in mitochondria as a membrane-integral long form and a short form that is soluble in the intermembrane space. Yeast strains that express temperature-sensitive mutants of Mgm1[9,10] or mammalian cells that lack OPA1 display fragmented mitochondria[11,12], which suggests that Mgm1 and OPA1 have an important role in inner-membrane fusion. Consistently, only the mitochondrial outer membrane—not the inner membrane—fuses in the absence of functional Mgm1[13]. Mgm1 and OPA1 have also been shown to maintain proper cristae architecture[10,14]; for example, OPA1 prevents the release of pro-apoptotic factors by tightening crista junctions[15]. Finally, the short form of OPA1 localizes to mitochondrial constriction sites, where it presumably promotes mitochondrial fission[16]. How Mgm1 and OPA1 perform their diverse functions in membrane fusion, scission and cristae organization is at present unknown. Here we present crystal and electron cryo-tomography structures of Mgm1 from *Chaetomium thermophilum*. Mgm1 consists of a GTPase (G) domain, a bundle signalling element domain, a stalk, and a paddle domain that contains a membrane-binding site. Biochemical and cell-based experiments demonstrate that the Mgm1 stalk mediates the assembly of bent tetramers into helical filaments. Electron cryo-tomography studies of Mgm1-decorated lipid tubes and fluorescence microscopy experiments on reconstituted membrane tubes indicate how the tetramers assemble on positively or negatively curved membranes. Our findings convey how Mgm1 and OPA1 filaments dynamically remodel the mitochondrial inner membrane.

We purified and crystallized a truncated short Mgm1 isoform from the thermophilic fungus *C. thermophilum* (hereafter denoted Mgm1) (Fig. 1a, Extended Data Fig. 1a, Supplementary Fig. 1). Crystals of this construct grown in the absence of nucleotides diffracted to a resolution of 3.6 Å. The structure was solved by single anomalous dispersion (Extended Data Fig. 1b, c, Extended Data Table 1).

Mgm1 contains four domains: a G domain, a bundle signalling element (BSE) domain, a stalk and a paddle domain (Fig. 1a, b). The G domain closely resembles that of human dynamin (Extended Data Fig. 2). An interface across the nucleotide-binding site (the 'G interface'), which is responsible for dimerization of G domains in the dynamin superfamily, is highly conserved in Mgm1 (Extended Data Fig. 1e). The adjacent BSE domain consists of three helices that are derived from different regions of Mgm1 (Fig. 1a, b). The BSE domain forms contacts with the G domain, as is the case in the closed

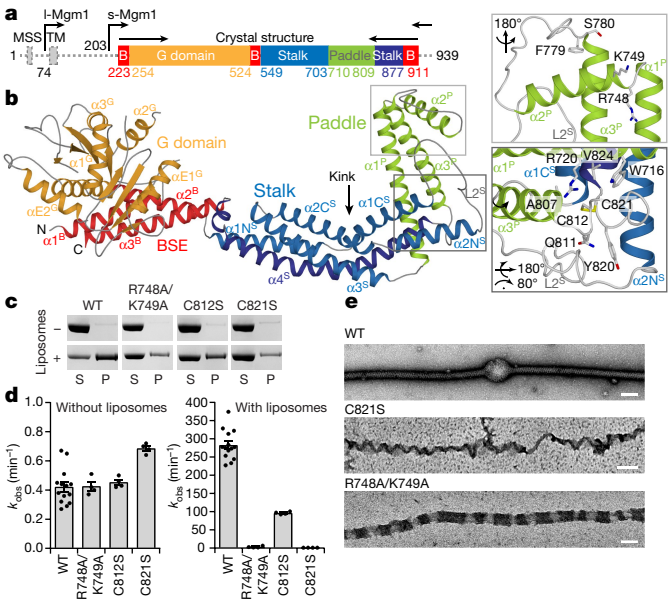


Fig. 1 | The structure of Mgm1 reveals a paddle domain that is required for membrane binding. a, Domain and isoform architecture of Mgm1. B, bundle signalling element; MSS, mitochondrial signal sequence; TM, transmembrane domain. l and s denote long and short isoforms of Mgm1, respectively. **b**, Ribbon representation of Mgm1. Domains are coloured individually as in **a**. The inset shows the disulfide bond between the conserved cysteine residues C812 and C821 and the conserved positively charged residues R748 and K749 in the paddle domain. Note that C821 is in the centre of the paddle whereas C812 is closer to its periphery. Apart from the loss of the disulfide bridge, the C821S mutation may therefore disrupt the paddle conformation more strongly than the C812S mutation. **c**, Representative liposome-binding experiment. P, pellet; S, supernatant; WT, wild type. $n = 4$ independent experiments. **d**, **e**, GTPase activity (**d**) ($n = 4$ independent experiments; data are mean ± s.e.m.) and negative-stain electron micrographs (**e**) of liposome tubulation of wild type Mgm1 and indicated mutants ($n = 2$ independent experiments; scale bars, 50 nm). Quantification of all experiments is shown in Extended Data Fig. 3a and the raw data is available in Supplementary Fig. 2.

[1]Crystallography, Max-Delbrück-Centrum for Molecular Medicine, Berlin, Germany. [2]Department of Structural Biology, Max Planck Institute of Biophysics, Frankfurt am Main, Germany. [3]Medical Biochemistry & Molecular Biology, Center for Molecular Signaling, PZMS, Saarland University Medical School, Homburg, Germany. [4]Biochemistry Department, University of Geneva, Geneva, Switzerland. [5]Alexander von Humboldt - Sofja Kovalevskaja Research Group, Max Planck Institute of Biophysics, Frankfurt am Main, Germany. [6]Institute of Biochemistry and Biotechnology, Section of Protein Biochemistry, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany. [7]EM facility, Max-Delbrück-Centrum for Molecular Medicine, Berlin, Germany. [8]Institute for Mathematics, Freie Universität Berlin, Berlin, Germany. [9]Institute of Chemistry and Biochemistry, Freie Universität Berlin, Berlin, Germany. [10]These authors contributed equally: Katja Faelber, Lea Dietrich. *e-mail: katja.faelber@mdc-berlin.de; werner.kuehlbrandt@biophys.mpg.de; oliver.daumke@mdc-berlin.de

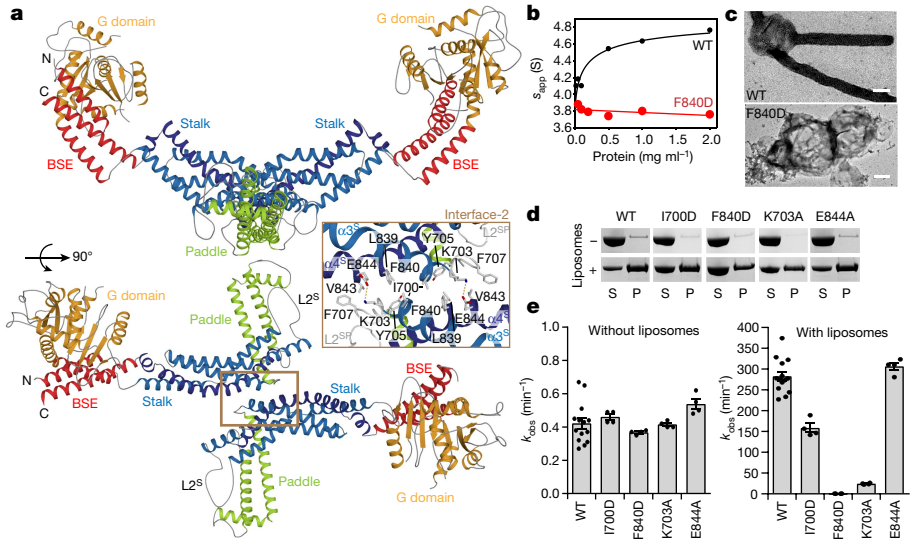


Fig. 2 | The Mgm1 dimer. a, The dimer is formed via interface-2 between opposing stalks. **b**, Sedimentation velocity analysis of Mgm1 and Mgm1(F840D) at different protein concentrations. The unusually small apparent sedimentation coefficients (s_{app}, expressed in Svedberg) of both the monomeric and the dimeric species are consistent with their non-globular structures. The data for Mgm1 can be fitted to a K_d of 1 μM. $n = 1$. **c**, Negative-stain electron micrographs of tubulated liposomes with wild-type Mgm1 and with Mgm1(F840D), in which an interface-2 residue is mutated. $n = 2$ independent experiments; scale bars, 50 nm. **d**, **e**, Liposome binding (**d**) and GTPase activity (**e**) for proteins with the stated interface-2 mutations ($n = 4$ independent experiments). For **e**, data are mean ± s.e.m.

conformation of dynamin[17–19]. The Mgm1 stalk domain is shorter than the dynamin stalk, comprising an antiparallel kinked four-helix bundle (Fig. 1b, Extended Data Fig. 2c, d). Unique to Mgm1 is the paddle domain, an elongated three-helix domain at the tip of the stalk, which is inserted between stalk helices $\alpha_3{}^S$ and $\alpha_4{}^S$ and contains a disulfide bridge that links Cys812 to Cys821 (Fig. 1b). Dynamin has a membrane-binding pleckstrin homology domain in the corresponding position.

We mutated two positively charged, surface-exposed residues as well as the two cysteines of the disulfide bridge in the paddle domain. In co-sedimentation experiments, the Mgm1 construct efficiently bound to Folch liposomes (lipids from bovine brain), whereas the mutant proteins bound less strongly (Fig. 1c, Extended Data Fig. 3a). Mgm1 bound to the non-hydrolysable GTP analogue GTPγS with a dissociation constant (K_d) of 9 μM (Extended Data Fig. 3b). The intrinsic GTPase activity of Mgm1 (about 0.5 min^{-1} at 37 °C) was accelerated about 500-fold in the presence of Folch liposomes, reaching rates of 270 min^{-1}. Yeast Mgm1 and human OPA1 also show increased GTPase activity in the presence of liposomes[20,21]. Stimulation of GTPase activity was considerably lower for all paddle mutants (Fig. 1d).

When incubated with liposomes, Mgm1 induced tubulation and coated the membrane surface in a regular pattern (Fig. 1e), with or without added nucleotide (Extended Data Fig. 3e, f). The membrane-remodelling activity of the paddle mutants was reduced, indicating that the paddle constitutes a membrane-binding site (Fig. 1e, Extended Data Fig. 3g).

The asymmetric unit of the crystals contained an Mgm1 dimer. The dimer interface (termed interface-2, in analogy to that of dynamin[17,18]) includes a hydrophobic core that is flanked by polar residues (Fig. 2a). Using analytical ultracentrifugation, we detected a concentration-dependent monomer–dimer equilibrium for Mgm1 (Fig. 2b, Extended Data Fig. 3c). The F840D mutation, in the centre of the hydrophobic dimer interface, rendered the protein monomeric. Assembly via the Mgm1 stalk interface-2 results in a V-shaped dimer, whereas dynamin stalks form an X-shaped dimer (Extended Data Fig. 2d, e).

Mutation of several interface-2 residues reduced both the binding of Mgm1 to liposomes and the extent to which its GTPase activity was stimulated in the presence of liposomes (Fig. 2d, e, Extended Data Fig. 3a). The F840D mutation had the most severe effect: Mgm1(F840D) failed to tubulate liposomes and did not form a regular protein pattern (Fig. 2c–e, Extended Data Fig. 3g). These results confirm the importance of interface-2 for the assembly of Mgm1 on the membrane surface.

We used a yeast model system to express Mgm1 mutants in a strain in which the expression of endogenous Mgm1 was under the control of the galactose-inducible and glucose-repressed *GAL1* promoter (Extended Data Fig. 4a). Loss of Mgm1 expression was associated with a rapid and irreversible loss of the mitochondrial genome, fragmentation of the mitochondrial network and the subsequent inability to switch to respiratory metabolism upon glucose depletion[6,10] (Extended Data Fig. 4b, c). Re-expression of yeast Mgm1 rescued the loss of mitochondrial respiratory function, as assessed by yeast growth, the presence of mitochondrially encoded cytochrome *c* oxidase 1 protein (Cox1) and restoration of the mitochondrial network (Extended Data Fig. 4b–g). Consistent with the liposome-binding assays, yeast Mgm1(F805D) (corresponding to F840D in *C. thermophilum* Mgm1)—but not Mgm1(N675A) (corresponding to I700D in *C. thermophilum* Mgm1, Supplementary Fig. 1)—failed to complement the loss of endogenous Mgm1 (Extended Data Fig. 4d, e, g). These results highlight the importance of interface-2 for Mgm1-dependent maintenance of mitochondrial DNA and respiration-competent mitochondria.

In the crystals, two Mgm1 dimers assembled into a tetramer via another stalk interface of approximately 1,000 $Å^2$; in further analogy to dynamin, we refer to this as stalk interface-1 (Fig. 3a). The tetramer is additionally stabilized by a contact of approximately 1,100 $Å^2$ between the BSE domain of one dimer and the stalk of the adjacent dimer. Interface-1 and the BSE–stalk contact site are highly conserved in the Mgm1 family (Extended Data Fig. 1e). Notably, the interface-1 interaction induces a bend of 20° between two stalk dimers (Extended Data Fig. 2e).

Mgm1(D559A) or Mgm1(K562A), containing mutations in interface-1, and Mgm1(Y537A) or Mgm1(R646A), containing mutations in the BSE–stalk contact, did not show major differences in liposome binding or GTPase activity compared to Mgm1 (Extended Data Fig. 3a, d). These mutant proteins also tubulated liposomes and formed a regular pattern on the membrane (Extended Data Fig. 3g). However, when these mutations were introduced into the corresponding positions of yeast Mgm1, the resultant mutant proteins failed to complement the loss of wild-type Mgm1 in respiratory growth (Fig. 3b), mitochondrial genome maintenance and mitochondrial morphology (Extended Data

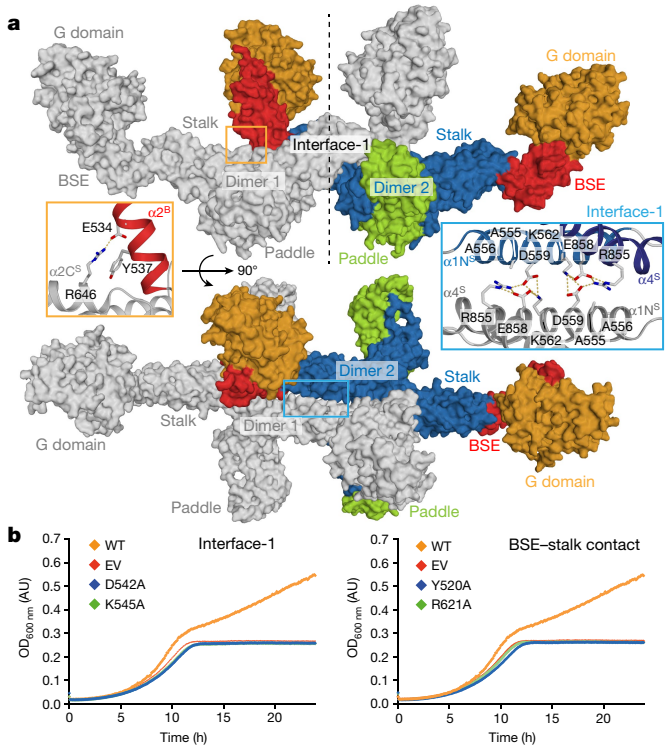


Fig. 3 | Assembly mechanism of Mgm1. a, The tetramer, as seen in the crystal. Two dimers (grey or colour-coded by domain) interact via stalk interface-1 and a small BSE–stalk contact. See insets for details. **b**, Yeast respiratory growth complementation assays with Mgm1 containing mutations in interface-1 residues (left; D542 and K545 in yeast Mgm1 correspond to D559 and K562 in the *C. thermophilum* protein) and residues of the BSE–stalk contact (right; Y520 and R621 in yeast Mgm1 correspond to Y537 and R646 in the *C. thermophilum* protein). A representative growth curve from $n = 3$ independent experiments is shown. EV, empty vector control; OD_{600nm}, optical density at 600 nm; AU, arbitrary units. See also Extended Data Fig. 4.

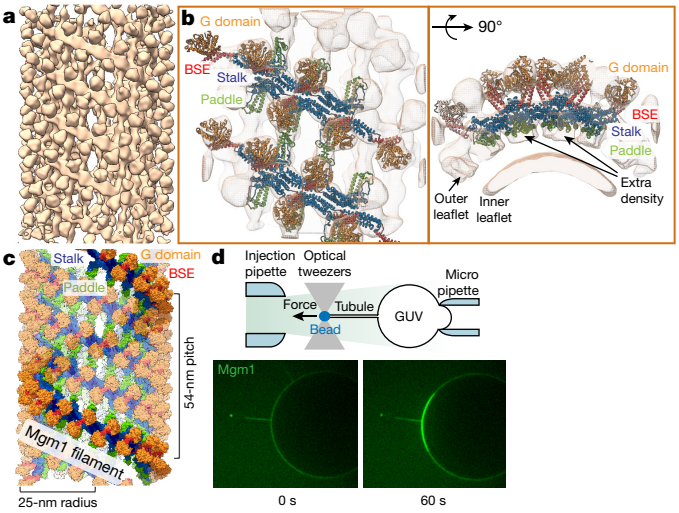


Fig. 4 | Mgm1 forms a helical lattice on the outside of lipid tubes. a, Mgm1 forms a regular protein coat on the outer surface of galactocerebroside-containing lipid tubes, enabling analysis by cryo-ET. **b**, The subtomogram average shows the protein lattice with Mgm1 flexibly fitted into the cryo-ET volume of the apo form. The outer leaflet of the membrane is not well defined in this reconstruction. Arrows indicate density not attributable to the protein, which was assigned to the outer membrane leaflet. **c**, Four filaments of Mgm1 dimers wrap around a membrane tube in a left-handed surface lattice. Stalks assemble via interface-1 and interface-2 in an alternating manner (Extended Data Fig. 7). **d**, Tube-pulling assay for the generation of a tube surface that is accessible from the outside. Mgm1 (green fluorescence) binds to the tube and the GUV surface ($n = 8$ independent experiments).

Fig. 4d, g). Notably, yeast Mgm1(Y520A)—containing a mutation in the BSE–stalk contact in the tetramer interface—exerted a strong dominant negative effect on respiratory yeast growth when co-expressed with endogenous Mgm1 (Extended Data Fig. 4h). The corresponding yeast strain retained mitochondrial DNA (Extended Data Fig. 4i), enabling us to examine Mgm1-specific deficits on mitochondrial morphology and ultrastructure. The expression of yeast Mgm1(Y520A) induced fragmentation of the mitochondrial network (Extended Data Fig. 4j, k), a reduction in the number and length of cristae and an increase in crista diameter (Extended Data Fig. 4l, m).

We used electron cryo-tomography (cryo-ET) and subtomogram averaging to determine the structure of membrane-bound Mgm1 (Fig. 4a, b, Extended Data Fig. 5a, b). In the absence of nucleotide, or upon the addition of GTPγS, Mgm1 remodelled Folch liposomes into membrane tubes of varying diameters ranging from around 18 nm to 140 nm. The Mgm1 coat decorated membrane tubes in a regular lattice. For subtomogram averaging, preformed tubes with diameters of around 20 nm were used in order to increase the number of particles for averaging. These tubes also stimulated the GTPase activity of Mgm1, although less strongly than did Folch liposomes (Extended Data Fig. 5c). The final resolution of the subtomogram average volume was 14.7 Å for the nucleotide-free and the nucleotide-bound forms (Extended Data Table 1). No substantial differences were apparent between the two volumes. Notably, the Mgm1 tetramer fits the subtomogram average volume with only minor positional domain rearrangements (Fig. 4b, Extended Data Fig. 6a, c, e). The G domain was in a closed conformation relative to the BSE domain and was located furthest from the membrane, the stalk was in the middle, and the paddle was next to the membrane. The Mgm1 coat in Fig. 4b, c can be viewed as a left-handed four-start helix, consisting of four parallel helical filaments (Extended Data Fig. 7a, b). Similar filaments were observed on Folch lipid tubes of different diameters, although their helical parameters varied (Extended Data Fig. 5d). The filament backbone was formed by stalks oligomerizing in an alternating fashion via interface-1 and interface-2. This is in contrast to dynamin filaments, in which the stalks oligomerize via three interfaces[22,23] (Extended Data Fig. 2e). Another difference compared with dynamin[22] is that we did not observe interactions between the G domains of adjacent helix turns. Instead, contact was established through the paddle domains (Fig. 4b, c, Extended Data Fig. 6a). Mutation of the conserved residues F779 and S780 in the paddle contact site affected membrane binding and stimulated GTPase activity only mildly (Extended Data Fig. 3a, d). Expression of the corresponding mutant protein in yeast complemented the loss of endogenous Mgm1 with respect to respiratory growth, but the cells exhibited moderate alterations of mitochondrial morphology and mitochondrial genome maintenance (Extended Data Fig. 4d, f, g).

The tendency for Mgm1 to form a left-handed helix on the convex exterior of membrane tubes is consistent with the curvature of the crystallographic tetramer that arises from the interaction between interfaces-1 of two dimers. A model in which several dimers are connected via identical interfaces-1 results in a continuous filament with dimensions and helix parameters (radius, pitch) similar to those observed by cryo-ET (Extended Data Fig. 7b, c). Microsecond-scale molecular dynamics simulations starting from the crystallographic tetramer provide further evidence for the curvature preference of the Mgm1 interface-1 (Extended Data Fig. 7d–k). The most likely curvature and twist in the simulated interface-1 was the same as that in the crystal lattice. The simulation results also suggested that there is sufficient flexibility in interface-1 to account for the observed variable radii of Mgm1 helical filaments.

We followed the dynamics of Mgm1 assembly on the membrane by live fluorescence confocal imaging. By manipulating streptavidin beads adhering to giant unilamellar vesicles (GUVs) with optical tweezers[24],

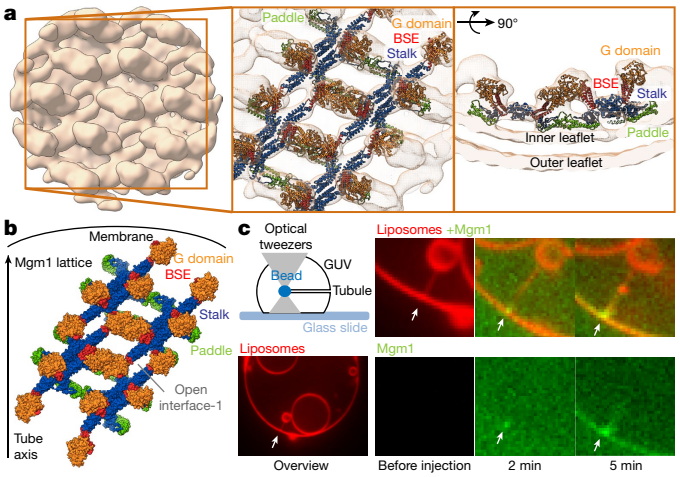


Fig. 5 | Mgm1 forms a lattice on the inside of lipid tubes. a, Mgm1 in the apo form decorates the inner surface of lipid tubes. The subtomogram average reveals regular protein arrays on the inner surface of membrane tubes; a magnified view shows Mgm1 flexibly fitted into the cryo-ET volume, with the G domains dimerized via the G interface. **b**, The Mgm1 lattice on the inner membrane surface of a tube. Interface-1 between tetramers is open, see also Extended Data Fig. 6. **c**, Tube-pulling assay for the generation of a membrane tube that is accessible from the inside. Red indicates lipid fluorescence. Mgm1 (green fluorescence) binds to the neck and the inner surface of the tube. The arrow points to the entry of the membrane tube ($n = 7$ independent experiments).

membrane tubes can be pulled out of the GUV in a controlled manner. Mgm1 was then injected into the chamber with a second pipette (Fig. 4d).

Consistent with the results of cryo-ET and negative-stain electron microscopy, Mgm1 adapted to different degrees of membrane curvature by decorating the outer surface of the membrane tube and the GUV. The addition of GTP after assembly did not result in membrane scission under these conditions, but the force required to hold the tube in place (measured as a function of bead displacement) increased by a factor of three to five. This is consistent with a GTP-dependent structural rearrangement of the Mgm1 coat and/or a GTP-dependent expansion of the membrane tube (Extended Data Fig. 8a, b).

Cryo-ET analysis revealed that Mgm1 occasionally decorated the inside surface of Folch membrane tubes in a regular pattern, which suggests that the liposomes were leaky (Fig. 5a, Extended Data Fig. 5e). In further experiments, liposomes were sonicated for a few seconds after the addition of Mgm1 to promote formation of the internal lattice. Tubes with an internal lattice were much wider and were less variable in diameter (range 90–105 nm) (Extended Data Fig. 5e). The negative (concave) membrane curvature on the inner surface of a larger tube resembles the inside of mitochondrial cristae. Subtomogram averages of Mgm1 decorating the inner vesicle surface were obtained for the nucleotide-free and the GTPγS-bound form (Fig. 5, Extended Data Fig. 5f, g, Extended Data Table 1). At an estimated resolution of 20.6 Å for the nucleotide-free and 18.8 Å for the nucleotide-bound form, the subtomogram average volumes appeared very similar.

As in the external lattice, the crystallographic Mgm1 tetramer fitted the subtomogram average volume of the internal lattice with only minor rearrangements (Fig. 5a, b, Extended Data Fig. 6b, d, e). The G domains were furthest from the membrane facing into the tube, the stalk was in the middle and the paddle domain was next to the membrane. The arrangement of tetramers on the internal lattice differed markedly from that on the external membrane surface (Fig. 5a, b, Extended Data Fig. 6a–e). Rather than through interface-1, assembly involved a contact between neighbouring tetramers that included conserved patches in the BSE and stalk domains, closely resembling the linear arrangement of tetramers in the crystal lattice (Extended Data Fig. 6f). The angle between filaments of Mgm1 tetramers and the plane perpendicular to the tube axis was 69°, whereas it was 21° in the external lattice (Fig. 5b, Extended Data Fig. 6a, b). Another major difference compared with the external lattice is that the G domains were in close contact, and their orientation indicated that interaction occurred through the G interface. This G domain contact was enabled by the opening of interface-1, even though the G domain/BSE interface remained closed. As with Mgm1 filaments on the surface of the outer membrane, the paddle domains contributed to lattice formation.

To investigate the assembly of Mgm1 on negatively curved membrane surfaces, streptavidin beads were pulled inside a GUV (Fig. 5c). In this situation, Mgm1 assembled preferentially at the funnel-shaped connection between the tube and the GUV and then grew further into the tube. Mgm1 did not redistribute on the membrane in the presence of GTP (Extended Data Fig. 8c). However, as with the positively curved (convex) membranes, the force on the tube increased in a GTP-dependent manner (Extended Data Fig. 8d). Together with the results from cryo-ET, the results of these experiments demonstrate conclusively that Mgm1 can form stable assemblies on negatively curved membranes.

Our study reveals the structural basis of Mgm1 assembly—via the stalks—into dimers, tetramers and helical filaments. Dynamin[17,18,22], dynamin-like MxA[25] and DNM1L[26,27] are likewise known to oligomerize via their stalks into helical filaments, although important parameters of the assembly mode differ (Extended Data Fig. 2e). In dynamin, the G domains of adjacent turns transiently dimerize and mediate a GTPase-dependent power stroke[19,28], which is thought to pull the filament turns against each other[29,30]. We propose that Mgm1 may undergo a similar power stroke, for the following reasons: the G domains and BSE domains of Mgm1 and dynamins are virtually identical; the mechanisms of membrane-stimulated GTPase activity are similar; the G domains in our cryo-ET reconstructions of Mgm1 are closely apposed; a GTP-dependent force was observed in the tube-pulling assays; and temperature-sensitive mutations in Mgm1 localize to one of the GTP-binding loops (switch I), the G/BSE domain interface or the assembly interface-1[10] (Extended Data Fig. 1d). Furthermore, the GTPase activity of OPA1 is required to sustain cristae morphology[14].

A dynamin-like power stroke would result in different remodelling processes depending on the assembly geometry of the Mgm1 filaments. When assembled on positively curved membranes in a left-handed helical pattern (Fig. 4), a dynamin-like power stroke would expand the diameter of the lipid tube. Conversely, a right-handed helix pattern would result in constriction, as observed in dynamin (Extended Data Fig. 9a, b, Supplementary Video 1).

Mgm1 is the only known member of the dynamin superfamily that can assemble on the inside of membrane tubes—a membrane geometry similar to that of mitochondrial inner membrane cristae. We postulate that Mgm1 can form helical filaments at the inside of membrane tubes with the shape and dimensions of crista junctions (Extended Data Fig. 9c, d). Notably, a power stroke in a left-handed helical assembly on the inside of a membrane tube would constrict its diameter—as observed for the crista junctions upon OPA1 overexpression[14]—whereas a right-handed assembly would expand it (Extended Data Fig. 9a, b, Supplementary Video 1). In Extended Data Fig. 9e–h, we suggest how the membrane geometry of different filament assemblies might explain inner-membrane fusion, scission, or the stabilization of cristae.

Taken together, our structural and functional studies reveal the molecular basis of the assembly of Mgm1 into filaments, and provide models of how the rearrangements of these filaments induce remodelling of the inner mitochondrial membrane.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1372-3.

Received: 4 February 2019; Accepted: 13 June 2019;
Published online 10 July 2019.

1. Nunnari, J. & Suomalainen, A. Mitochondria: in sickness and in health. *Cell* **148**, 1145–1159 (2012).
2. Youle, R. J. & van der Bliek, A. M. Mitochondrial fission, fusion, and stress. *Science* **337**, 1062–1065 (2012).
3. van der Laan, M., Horvath, S. E. & Pfanner, N. Mitochondrial contact site and cristae organizing system. *Curr. Opin. Cell Biol.* **41**, 33–42 (2016).
4. Pernas, L. & Scorrano, L. Mito-morphosis: mitochondrial fusion, fission, and cristae remodeling as key mediators of cellular function. *Annu. Rev. Physiol.* **78**, 505–531 (2016).
5. Wai, T. & Langer, T. Mitochondrial dynamics and metabolic regulation. *Trends Endocrinol. Metab.* **27**, 105–117 (2016).
6. Jones, B. A. & Fangman, W. L. Mitochondrial DNA maintenance in yeast requires a protein containing a region related to the GTP-binding domain of dynamin. *Genes Dev.* **6**, 380–389 (1992).
7. Alexander, C. et al. OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. *Nat. Genet.* **26**, 211–215 (2000).
8. Delettre, C. et al. Nuclear gene OPA1, encoding a mitochondrial dynamin-related protein, is mutated in dominant optic atrophy. *Nat. Genet.* **26**, 207–210 (2000).
9. Wong, E. D. et al. The dynamin-related GTPase, Mgm1p, is an intermembrane space protein required for maintenance of fusion competent mitochondria. *J. Cell Biol.* **151**, 341–352 (2000).
10. Meeusen, S. et al. Mitochondrial inner-membrane fusion and crista maintenance requires the dynamin-related GTPase Mgm1. *Cell* **127**, 383–395 (2006).
11. Cipolat, S., Martins de Brito, O., Dal Zilio, B. & Scorrano, L. OPA1 requires mitofusin 1 to promote mitochondrial fusion. *Proc. Natl Acad. Sci. USA* **101**, 15927–15932 (2004).
12. Ishihara, N., Fujita, Y., Oka, T. & Mihara, K. Regulation of mitochondrial morphology through proteolytic cleavage of OPA1. *EMBO J.* **25**, 2966–2977 (2006).
13. Meeusen, S., McCaffery, J. M. & Nunnari, J. Mitochondrial fusion intermediates revealed *in vitro*. *Science* **305**, 1747–1752 (2004).
14. Frezza, C. et al. OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. *Cell* **126**, 177–189 (2006).
15. Yamaguchi, R. et al. Opa1-mediated cristae opening is Bax/Bak and BH3 dependent, required for apoptosis, and independent of Bak oligomerization. *Mol. Cell* **31**, 557–569 (2008).
16. Anand, R. et al. The i-AAA protease YME1L and OMA1 cleave OPA1 to balance mitochondrial fusion and fission. *J. Cell Biol.* **204**, 919–929 (2014).
17. Faelber, K. et al. Crystal structure of nucleotide-free dynamin. *Nature* **477**, 556–560 (2011).
18. Ford, M. G., Jenni, S. & Nunnari, J. The crystal structure of dynamin. *Nature* **477**, 561–566 (2011).
19. Chappie, J. S., Acharya, S., Leonard, M., Schmid, S. L. & Dyda, F. G domain dimerization controls dynamin's assembly-stimulated GTPase activity. *Nature* **465**, 435–440 (2010).
20. Ingerman, E. et al. Dnm1 forms spirals that are structurally tailored to fit mitochondria. *J. Cell Biol.* **170**, 1021–1027 (2005).
21. Ban, T., Heymann, J. A., Song, Z., Hinshaw, J. E. & Chan, D. C. OPA1 disease alleles causing dominant optic atrophy have defects in cardiolipin-stimulated GTP hydrolysis and membrane tubulation. *Hum. Mol. Genet.* **19**, 2113–2122 (2010).
22. Kong, L. et al. Cryo-EM of the dynamin polymer assembled on lipid membrane. *Nature* **560**, 258–262 (2018).
23. Reubold, T. F. et al. Crystal structure of the dynamin tetramer. *Nature* **525**, 404–408 (2015).
24. Chiaruttini, N. et al. Relaxation of loaded ESCRT-III spiral springs drives membrane deformation. *Cell* **163**, 866–879 (2015).
25. Gao, S. et al. Structure of myxovirus resistance protein a reveals intra- and intermolecular domain interactions required for the antiviral function. *Immunity* **35**, 514–525 (2011).
26. Frohlich, C. et al. Structural insights into oligomerization and mitochondrial remodelling of dynamin 1-like protein. *EMBO J* **32**, 1280–1292 (2013).
27. Kalia, R. et al. Structural basis of mitochondrial receptor binding and constriction by DRP1. *Nature* **558**, 401–405 (2018).
28. Chappie, J. S. et al. A pseudoatomic model of the dynamin polymer identifies a hydrolysis-dependent powerstroke. *Cell* **147**, 209–222 (2011).
29. Roux, A., Uyhazi, K., Frost, A. & De Camilli, P. GTP-dependent twisting of dynamin implicates constriction and tension in membrane fission. *Nature* **441**, 528–531 (2006).
30. Antonny, B. et al. Membrane fission by dynamin: what we know and what we need to know. *EMBO J.* **35**, 2270–2284 (2016).
31. Doublié, S. Preparation of selenomethionyl proteins for phase determination. *Methods Enzymol.* **276**, 523–530 (1997).
32. Kabsch, W. XDS. *Acta Cryst. D* **66**, 125–132 (2010).
33. Sparta, K. M., Krug, M., Heinemann, U., Mueller, U. & Weiss, M. S. Xdsapp2.0. *J. Appl. Crystallogr.* **49**, 1085–1092 (2016).
34. Terwilliger, T. C. et al. Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallogr. D* **65**, 582–601 (2009).
35. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
36. Echols, N. et al. Graphical tools for macromolecular crystallography in PHENIX. *J. Appl. Crystallogr.* **45**, 581–586 (2012).
37. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
38. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
39. Sievers, F. & Higgins, D. G. Clustal omega. *Curr. Protoc. Bioinform.* **48**, 1.25.1–1.25.33 (2014).
40. Schuck, P. Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
41. Longtine, M. S. et al. Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
42. Yofe, I. & Schuldiner, M. Primers-4-Yeast: a comprehensive web tool for planning primers for *Saccharomyces cerevisiae*. *Yeast* **31**, 77–80 (2014).
43. Sikorski, R. S. & Hieter, P. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics* **122**, 19–27 (1989).
44. Ieva, R. et al. Mgr2 functions as lateral gatekeeper for preprotein sorting in the mitochondrial inner membrane. *Mol. Cell* **56**, 641–652 (2014).
45. Morgenstern, M. et al. Definition of a high-confidence mitochondrial proteome at quantitative scale. *Cell Rep.* **19**, 2836–2852 (2017).
46. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
47. Wilson-Kubalek, E. M., Brown, R. E., Celia, H. & Milligan, R. A. Lipid nanotubes as substrates for helical crystallization of macromolecules. *Proc. Natl Acad. Sci. USA* **95**, 8040–8045 (1998).
48. Hagen, W. J. H., Wan, W. & Briggs, J. A. G. Implementation of a cryo-electron tomography tilt-scheme optimized for high resolution subtomogram averaging. *J. Struct. Biol.* **197**, 191–198 (2017).
49. Mastronarde, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
50. Grant, T. & Grigorieff, N. Measuring the optimal exposure for single particle cryo-EM using a 2.6 Å reconstruction of rotavirus VP6. *eLife* **4**, e06980 (2015).
51. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
52. Castaño-Díez, D., Kudryashev, M., Arheit, M. & Stahlberg, H. Dynamo: a flexible, user-friendly development tool for subtomogram averaging of cryo-EM data in high-performance computing environments. *J. Struct. Biol.* **178**, 139–151 (2012).
53. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
54. Whitford, P. C. et al. Excited states of ribosome translocation revealed through integrative molecular modeling. *Proc. Natl Acad. Sci. USA* **108**, 18943–18948 (2011).
55. Noel, J. K. et al. SMOG 2: a versatile software package for generating structure-based models. *PLOS Comput. Biol.* **12**, e1004794 (2016).
56. Harvey, M. J. & De Fabritiis, G. AceCloud: molecular dynamics simulations in the cloud. *J. Chem. Inf. Model.* **55**, 909–914 (2015).
57. Best, R. B. et al. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone φ, ψ and side-chain χ_1 and χ_2 dihedral angles. *J. Chem. Theory Comput.* **8**, 3257–3273 (2012).
58. Pronk, S. et al. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
59. Theile, C. S. et al. Site-specific N-terminal labeling of proteins using sortase-mediated reactions. *Nat. Protoc.* **8**, 1800–1807 (2013).

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and purification. *C. thermophilum* Mgm1 (Mgm1, amino acids 219–912) and indicated mutants of this construct were expressed from pET46-EK/LIC (Novagen) as N-terminal His_6-tag fusion followed by a PreScission cleavage site. Proteins were expressed in *Escherichia coli* host strain BL21-DE3, and bacteria were cultured in TB medium at 37 °C followed by induction with 200 μM isopropyl-β-D-thiogalactopyranoside and a temperature shift to 20 °C for overnight expression. Selenomethionine-substituted Mgm1 was expressed in M9 minimal medium, supplemented with L-amino acids lysine, phenylalanine, threonine (100 mg l^{-1}), isoleucine, leucine, valine and selenomethionine (50 mg l^{-1}), using the same vector and host strain as for native protein expression[31]. Cells were resuspended in buffer A (25 mM HEPES/NaOH (pH 7.8), 350 mM NaCl, 150 mM KCl, 2 mM $MgCl_2$, 1 μM DNase (Roche), 500 μM Pefabloc (Roth)) and disrupted by a microfluidizer (Microfluidics). Cleared lysates (95,000*g*, 1 h, 4 °C) were incubated with Benzonase (Novagen) for at least 30 min at 4 °C before application to a Co^{2+}-Talon column (Clontech). Proteins were eluted with buffer A containing an additional 100 mM imidazole. Fractions containing Mgm1 were incubated with 2.4 mM β-mercaptoethanol and His_6-tagged Prescission protease overnight at 4 °C. Then, imidazole, β-mercaptoethanol and the free His-tag were removed by using 50 kDa molecular weight cut-off concentrators (Amicon) and washing with buffer A, before a second application to a Co^{2+}-Talon column to remove non-cleaved His-tagged Mgm1 and protease. The flow-through and four column volumes of washing buffer A were collected and concentrated. Finally, Mgm1 was purified by size-exclusion chromatography on a Superdex200 column (GE) in buffer A. Fractions containing Mgm1 were pooled, concentrated and flash-frozen in liquid nitrogen (Extended Data Fig. 1a). Selenomethionine-substituted protein and mutant proteins were purified using the same protocol.

Crystallization and structure determination. Crystallization trials using the sitting-drop vapour-diffusion method were performed at 4 °C using a Gryphon pipetting robot (Art Robbins Instruments) and Rock Imager storage system (Formulatrix). 300 nl of the selenomethionine-substituted Mgm1 at a concentration of 12.9 mg ml^{-1} was mixed with an equal volume of reservoir solution containing 8% PEG400, 3% isopropanol, 100 mM Na-citrate buffer (pH 5.5). Crystals appeared after 2 weeks and had final dimensions of 500 μm × 200 μm × 50 μm. During flash-cooling of the crystals in liquid nitrogen, a cryo-solution containing additionally 20% ethylene glycol was used. The dataset was recorded at BL14.1 at BESSY II, Berlin. One native dataset was collected at a wavelength of 0.9794 Å and a temperature of 100 K from a single crystal and processed and scaled using the XDS program suite[32,33]. Twenty-two out of twenty-six Se sites were detected by Autosol/PHENIX[34] for two molecules in the asymmetric unit (80% solvent content). The density showed a continuous trace for the peptide backbone and clear anomalous signals for the positions of the selenomethionine side chains. The initial model was built by adapting the BSE and stalk domain from the human dynamin 3 structure (Protein Data Bank (PDB): 5A3F) to the density. For chain A, the nucleotide-free G domain of human dynamin 1 (PDB: 2AKA) fitted the density well, whereas density for the G domain of chain B was weak. The G domain was therefore omitted in the initial chain B model. The electron density for loop 2 in the stalk ($L2^S$, see Supplementary Fig. 1) and for the paddle was well defined. Building of missing residues in this area was guided by the anomalous signal of the selenomethionine side chains. The model was built using COOT[35] and iteratively refined with Phenix 1.11.1-2575[36] including Hendrickson–Lattman coefficients, non-crystallographic symmetries of the respective domains, secondary structure restraints, one TLS group per domain and one B factor per amino acid. Occupancy of side chains with considerable radiation damage was reduced to 0.8 or 0.6 for surface exposed glutamate or aspartate residues, and to 0.8 or 0.5 for selenomethionine residues. Finally, the G domain from chain A was transplanted to chain B and refined as a rigid body. Two ethylene glycol molecules were built into remaining difference density at the end of the refinement. 1,252 residues out of 1,304 refined residues (96%) are in the most favoured regions of the Ramachandran plot and 3 residues are in the disallowed regions (0.23%), as analysed with Phenix. Buried surface areas were calculated using the PISA server[37]. Domain superpositions were performed with lsqkab from the CCP4 program suite[38]. Figures were prepared with the PyMOL Molecular Graphics System, Version 2.0 (Schrödinger, LLC.). Sequences were aligned using CLUSTAL Omega[39] and adjusted by hand.

Analytical ultracentrifugation experiments. All measurements were performed in 25 mM HEPES/NaOH pH 7.8, 50 mM NaCl, 150 mM KCl and 1 mM $MgCl_2$ at 20 °C using an Optima XL-A centrifuge (Beckman) and an An50Ti rotor equipped with double sector cells. Depending on protein concentration, the distribution of the protein in the cell was monitored at 230 or 280 nm. Data were analysed using the software SedFit[40]. Sedimentation velocity was run at 40,000 r.p.m. for 3 h, sedimentation equilibrium was performed at 8,000 r.p.m.

Liposome co-sedimentation assays. Liposomes were prepared as previously described (www. endocytosis.org). Folch liposomes (0.6 mg ml^{-1}) (total bovine brain lipids fraction I from Sigma) in 25 mM HEPES/NaOH (pH 7.8), 60 mM NaCl, 100 mM KCl and 0.5 mM $MgCl_2$ were incubated at room temperature with 4 μM of the indicated Mgm1 construct for 10 min in 40 μl reaction volume, followed by spinning at 210,000*g* for 10 min at 20 °C and SDS–PAGE analysis of the supernatant and the pellet. For quantification, the protein bands were integrated using ImageJ and the intensity of each band (supernatant or pellet) was divided by the sum of the intensities from supernatant and pellet.

Isothermal titration calorimetry. Isothermal titration calorimetry experiments were performed at 18 °C in a PEAQ-ITC (Malvern) in 20 mM HEPES/NaOH pH 7.5, 60 mM NaCl, 100 mM KCl and 0.5 mM $MgCl_2$, with 50 μM Mgm1 in the reaction chamber and 1 mM GTPγS in the syringe. Malvern software was used to integrate the binding isotherms and calculate the binding parameters.

GTP hydrolysis assay. GTPase activities of 1 μM of the indicated Mgm1 constructs were determined at 37 °C in 25 mM HEPES/NaOH (pH 7.8), 60 mM NaCl, 100 mM KCl and 0.5 mM $MgCl_2$, in the absence and presence of 0.1 mg ml^{-1} Folch liposomes, using saturating concentrations of GTP as substrate (1 mM for the basal and 3 mM for the stimulated reactions). Reactions were initiated by the addition of protein to the reaction. At different time points, reaction aliquots were diluted 15-fold with GTPase buffer and quickly frozen in liquid nitrogen. Samples were analysed with an HPLC system (Agilent Technologies). Denatured proteins were adsorbed to a C18 guard column and nucleotides were separated via a reversed-phase Hypersil ODS-2 C18 column (250 × 4 mm), with 10 mM tetrabutylammonium bromide, 100 mM potassium phosphate (pH 6.5), 7.5% acetonitrile as running buffer. Nucleotides were detected by absorption at 254 nm and quantified by integration of the corresponding peaks. Rates were derived from a linear fit to the initial reaction.

Negative-stain electron microscopy. For electron microscopy of negatively stained samples in a Zeiss EM910, 4 μM Mgm1 (amino acids 219–912) in 25 mM HEPES/NaOH (pH 7.8), 60 mM NaCl, 100 mM KCl, 0.5 mM $MgCl_2$ and 3 mM guanosine-5′-[(β,γ)-methyleno]triphosphate were incubated at room temperature for 10 min. The final concentration of unfiltered Folch liposomes was 0.6 mg ml^{-1}. Samples were incubated on carbon-coated copper grids (Plano) and stained with 2% uranyl acetate.

Yeast growth assay. To test the ability of mutant Mgm1 variants to complement the loss of wild-type Mgm1 in yeast (*Saccharomyces cerevisiae*), a *GAL1* promoter was inserted upstream of the *MGM1* open reading frame by homologous recombination. To this end, the *GAL1* promoter was amplified from pFA6a-kanMX6-PGAL1[41] (using oligonucleotides MGM1-PGAL-FW CATCCCAAGAGTGGCGAACTATAACACATTAGTA AGGATGgaattcgagctcgtttaaac and MGM1-PGAL-REV GCTGTCTT CTCAGAATTAAAAGCCGTACTGGGCTCGCATTcattttgagatccgggtttt[42]) and transformed into the YPH499 wild-type strain[43]. Mutations were introduced into pRS414-Mgm1[44] by site-directed mutagenesis. The *PGAL1-MGM1* yeast strain was transformed with the empty vector pRS414 or pRS414-Mgm1 encoding wild-type Mgm1 or mutant variants. After selection on synthetic defined (-TRP) media (0.67% (w/v) YNB without amino acids (BD Difco), -TRP amino acid drop-out mix (MP Biomedicals)) containing 2% (w/v) galactose and 1% (w/v) raffinose, yeast were grown in media containing 2% (w/v) glucose as a carbon source to suppress expression of the endogenous wild-type Mgm1 allele. Under these conditions, cells expressing no or non-functional Mgm1 rapidly lose mitochondrial DNA[6]. Subsequently, cultures were diluted in media containing 0.2% (w/v) glucose in 48-well microtiter plates and growth was monitored for 24 h at 30 °C using a Tecan Spark 10M microplate reader by measuring the absorbance at 600 nm every 5 min after a 10 s linear shake with an amplitude of 2.5 mm at 630 r.p.m. Between cycles, the plate was agitated in a double-orbital manner with an amplitude of 1.5 mm at 180 r.p.m. Blank-corrected mean absorbance values from two or three wells per mutant strain were plotted using GraphPad Prism 6.0 and growth experiments were repeated with cell populations from three independent yeast transformations.

To test for dominant-negative effects of Mgm1 mutants, the wild-type strain YPH499 was transformed with pRS414-Mgm1 encoding wild-type or mutant Mgm1 and growth was assessed in synthetic defined medium containing 3% (v/v) glycerol as described above. To test whether Mgm1 variants are stably expressed in yeast cells and able to retain mitochondrial DNA, mitochondria were isolated on a small scale[45] and analysed by SDS–PAGE and western blotting using antibodies directed against Mgm1, Cox1 (mitochondrially encoded cytochrome *c* oxidase subunit 1) and Ssc1 (mitochondrial Hsp70, loading control).

Yeast microscopy. Yeast cells were grown in synthetic defined (-TRP) media containing either 2% (w/v) glucose (for *PGAL1-MGM1* yeast strains expressing plasmid-borne Mgm1 variants) or 3% (v/v) glycerol (for dominant-negative mutant strains) to mid-logarithmic phase and stained with 0.5 μg ml^{-1} DAPI (4′,6-diamidino-2-phenylindole) and 175 nM $DiOC_6$ (3,3′-dihexyloxacarbocyanine iodide) in 5% (w/v) glucose and 10 mM HEPES (pH 7.2). Immediately

after staining, *Z*-stacks were recorded on a Leica DMi8 fluorescent microscope with a 63×/1.40 objective and a Leica DFC3000 G CCD camera. Images were deconvoluted with Huygens Essential (Scientific Volume Imaging, http://svi.nl) and maximum intensity projections were created in Fiji[46]. Contrast was adjusted linearly to correct for variations in $DiOC_6$ uptake. For quantification of mitochondrial morphology, cells with tubular or fragmented mitochondrial networks were counted in images from three independent cultures (for each culture at least 70 cells were counted). Wild-type ρ^0 cells were generated by ethidium bromide treatment of ρ^+ cells.

Electron microscopy of yeast mitochondria. Yeast cells were fixed for 3 h with 4% (w/v) paraformaldehyde and 0.5% (v/v) glutaraldehyde in 0.1 M citrate buffer (pH and temperature adjusted to growth conditions). Samples were treated with 1% (w/v) sodium metaperiodate for 1 h at room temperature. Yeast cells were embedded in 10% (w/v) gelatin, infiltrated with 2.3 M sucrose and frozen in liquid nitrogen. Ultrathin sections were cut at −115 °C (Reichert Ultracut S, Leica) and collected on 200-mesh copper grids (Plano) coated with Formvar and carbon. Sections were stained with 3% (w/v) tungstosilicic acid hydrate in 2.5% (w/v) polyvinyl alcohol. Samples were examined at 80 kV with a Zeiss EM 910 electron microscope (Zeiss), and images were recorded with a Quemesa CCD camera and the iTEM software (Emsis). Images were analysed by ImageJ/Fiji[46]. All applied statistical tests were calculated using Prism (GraphPad software). A normality distribution test (Kolmogorov–Smirnov test) was carried out for all experimental values, and with normally distributed data a Student's *t*-test (two-tailed *P* value) was applied, otherwise the Mann–Whitney rank-sum (two-tailed *P* value) test was used to calculate the significant difference between two groups.

Liposome preparation for cryo-ET. For examining Mgm1 assembly on membranes by electron cryo-tomography, dried lipids were rehydrated to a final concentration of 3 mg ml^{-1} in liposome buffer (20 mM HEPES, pH 7.5, 150 mM NaCl). Folch lipids (brain extract from bovine brain, type I, fraction I, Sigma-Aldrich) were used for inside decoration, or a lipid mixture of 70% galactocerebroside[47], 10% cardiolipin (both Sigma-Aldrich) and 20% di-oleyl-phosphatidylcholine (DOPC) (Avanti Polar Lipids) for outside decoration of tubes. Liposomes were prepared by sonication followed by extrusion through a 1 μm polycarbonate filter. Rehydrated lipids were incubated with purified Mgm1 (final concentration 10 μM) for 30 min at room temperature in the absence or presence of GTPγS (final concentration 1 mM, Jena Bioscience). For inside decoration, Mgm1 (with or without nucleotide) was added before the liposome preparation step.

Grid preparation and image acquisition for electron cryo-tomography. The final sample was mixed in a 1:1 ratio with colloidal gold fiducial markers and 3 μl were applied to freshly glow-discharged R2/2 Cu 300-mesh holey carbon-coated support grids (Quantifoil Micro Tools). Grids were plunge-frozen using a Vitrobot Mark IV plunge-freezer at 100% humidity and 10 °C. Samples were imaged in a FEI Titan Krios electron microscope (FEI Company) operating at 300 kV, equipped with a K2 summit direct electron detector and Quantum energy filter (Gatan). The nominal magnification was set to 53,000×, yielding a calibrated pixel size of 2.7 Å. Tomographic tilt series were acquired following a dose-symmetric tilting scheme[48] with a 3° increment and a cumulative total electron dose of approximately 90 e^- $Å^{-2}$. Defocus values ranged from −2.0 to −4.0 μm. Data were acquired with the SerialEM software package[49] in dose-fractionation mode.

Tomogram reconstruction and subtomogram averaging. Dose-fractionated movies of tomograms were aligned using either Unblur[50] or MotionCor2[51]. After contrast-transfer-function correction, images were combined to generate a raw image stack that was used as input for generating tomograms with IMOD. Single tilt-images were aligned by gold fiducial markers and volumes reconstructed by weighted back-projection. Particle extraction, alignment and subtomogram averaging were performed with Dynamo[52] and MATLAB. For a whole tube, particles were picked along the filaments using the respective option in the Dynamo toolbox. Eighteen membrane tubes covered with a clear visible protein coat within 15 different tomograms and 10 tubes within 10 tomograms were used for processing for the apo form and the GTPγS bound form, respectively. Owing to the differences in diameter of the inside decoration, only two membrane tubes in two individual tomograms were used for the apo form as well as the GTPγS bound state. For close-up views, tubes were sub-boxed along the helical pattern. For tubes decorated on the inside, particles were picked along the wall of the lipid tube. Before subtomogram averaging, the datasets were divided into two independent half sets for resolution estimation. Each half set was aligned to an independent reference generated from a subset of each half set and reference-free alignment. To address the possibility of different handedness, classification was performed during the processing workflow. Only protein assemblies with a left-handed helical pattern were observed. To exclude that the left-handed arrangement of the outside decoration was driven by the preformed lipid tubes, subtomogram averaging of Mgm1 covering the outside of Folch lipid tubes of different diameters was performed. Also in this case, only protein assemblies with a left-handed helical pattern were observed. The numbers of particles that contributed to the converged averages of the main structures and final resolution from Fourier shell correlation (FSC) curves are listed in Extended Data Table 2. The final structures were obtained using relion_reconstruct from the Relion toolbox. USCF Chimera and MATLAB were used for structure and FSC curve display, respectively[53].

Molecular dynamics simulations. *Flexible fitting into cryo-ET volume.* A general approach for building atomic models from cryo-ET reconstructions is to include a potential energy term coupling the atomic coordinates during a molecular simulation to the experimentally determined density. Here we used the MDfit method[54], which uses an all-atom structure-based model[55] based on the tetramer crystal structure, and additionally includes an energetic term that attempts to maximize the correlation between the experimental density and the simulated density of the molecular dynamics trajectory. The structure-based model has an explicit energy minimum at the tetramer crystal structure, which means that the secondary structure seen in the crystal is maintained during flexible fitting. Modified Gromacs source code containing MDfit and software for creating the all-atom structure-based topologies are available for download at[55] http://smog-server.org. Default MDfit parameters were used, including setting the energetic weight of the map equal to the number of atoms. For both the inner and outer decoration, the initial configuration was generated by manually placing twelve tetramers (247,728 heavy atoms) into and surrounding the cryo-ET volume with the aid of the "Fit in Map" tool in Chimera[53]. Simulations were performed until the cross-correlation stabilized. Only the dimers that were completely within the cryo-ET volume were saved for deposition alongside the cryo-ET volume. After fitting the inner decoration, G domains appeared to be in contact. This was checked by strongly constraining the G domains to form the G interface (as in dynamin). The fit including the constraint was nearly identical to that without, suggesting that the cryo-ET for the inner decoration contains the canonical G interface. The submitted model includes the constraint.

All-atom molecular dynamics to support the "pre-shaped" tetramer and characterize its flexibility. A 2.6-μs all-atom molecular dynamics simulation of a stalk tetramer in explicit solvent was performed to estimate its shape in the absence of crystal interactions. The simulation was initialized from the tetramer crystal structure with a closed interface-1 and contained for each monomer the four stalk helices (residues 549–590, 635–720 and 828–877). Two G–G–S–G–G linkers were used to connect breaks in the stalk where the paddle was cut out, creating a single chain for each stalk monomer. Simulations were performed with Acellera ACEMD[56] using the CHARMM36 forcefield[57]. Details of the simulation are as follows: NPT ensemble, temperature 300K, Langevin thermostat, Berendsen barostat at 1 atm, restrained bonds, timestep 4 fs, PME electrostatics, grid spacing 1 Å, cutoff 9 Å, switching at 7.5 Å. The conformation of the stalk tetramer was analysed to estimate the structural preference and flexibility of a stalk filament containing the tetramer. See Extended Data Fig. 7 for details.

All-atom structure-based model for inner decoration of 1-start helix. Our aim was to determine the tetramer structure upon confinement in a filament decorating the interior of a narrow membrane tube ($r = 30$ nm) with a small pitch ($P = 12$ nm). In particular, we were interested in whether the crystallographic interfaces-1 and -2 can be consistent with negatively curved geometries. To this end, a molecular dynamics simulation was performed on a short filament (octamer) using a simplified potential that includes the all-atom geometry. Three constraints were imposed: the putative membrane-binding residues R748 and K749 in each monomer were constrained to a 30 nm radius from the *z*-axis; an impenetrable cylindrical wall was imposed with a 30 nm radius; and the *z* coordinate of the centres of mass of each dimer (interface-2) were constrained such that the short filament had an effective pitch of 12 nm. No restraints were introduced in interface-1. The simulation potential was an all-atom structure-based model using the tetramer crystal structure with interface-1 formed. The simulation topology for Gromacs[58] was created using the tetramer crystal and SMOG2.1 with the default forcefield 'SBM_AA'[55]. The octamer topology was created by merging two tetramer topologies and additionally copying the requisite pair interactions for the new interface-1 created by connecting the tetramers. Langevin dynamics with a low temperature (0.16 reduced units, 20K Gromacs temperature) for 10×10^6 steps was used to get near to the minimum energy subject to the constraints. A steepest-descent minimization was used for the final analysed configuration. To minimize edge effects, the interior tetramer of the octamer filament was analysed.

Tube-pulling assays. Mgm1 was labelled with a fluorescein-labelled peptide using a sortase-mediated reaction[59]. All lipids were purchased from Avanti Polar Lipids. GUVs were electroformed[24] from a lipid mix (2 mg ml^{-1}) containing DOPC, di-oleyl-phosphatidylserine, rhodamine-phosphatidylethanolamine (Rhod-PE) and di-sialyl-phosphatidylserine-polyethylene-glycol-2000-biotin (DSPE-PEG(2000)Biotin), at a ratio of 7:3:0.01:0.003. GUVs were then transferred to a microscopy chamber of two rectangular glass slides (11 × 35 mm) and mounted on an inverted microscope including a Nikon Eclipse Ti base, a CSU-X1 confocal system (Nikon), an Andor Ixon EMCCD camera (Oxford Instruments) and homemade optical tweezers consisting of a 5-W, 1,064-nm

laser (ML5-CW-P-TKS-OTS, Manlight) focused through a 100 × 1.3 numerical aperture oil objective. Images were acquired using SlideBook software (Intelligent Imaging Innovation). Bead traces were acquired with a C-MOS Camera (Picelink) using custom-made software. Outward membrane nanotubes were formed by holding a 3.05-μm streptavidin-coated polystyrene bead (Spherotech) glued onto a GUVs with optical tweezers, while pulling away the GUVs held by aspiration with a hand micropipette and controlled with motorized micromanipulators (MP-285, Sutter Instrument). Subsequently, Mgm1 was diluted to a final concentration of 3 μM in 20 mM HEPES/NaOH pH 7.4, 200 mM NaCl and 1 mM $MgCl_2$, and injected in the vicinity of the membrane tube using a second micropipette connected to a pressure control system (MFCS-VAC, −69 mbar, Fluigent). For pulling membrane nanotubes inward, 2.01 μm glass beads (Bangs Laboratories) were internalized with optical tweezers into GUVs adhering to an Avidin-coated flow chamber (coverslip and sticky-Slide VI 0.4, Ibidi). Tubes were pulled by moving the stage, and thus the GUV. Mgm1 (3 μM) was added with a syringe pump (Aladdin, World Precision Instruments) connected to the Ibidi flow chamber. The force F was determined by applying Hooke's law $F = k\Delta x$ to the bead displacement Δx and trap stiffness k (3.05 μm beads: $k = 79$ pN nm^{-1}; 2.01 μm beads: $k = 75$ pN nm^{-1}). The basis of inward-pulled tubes was unstable and moved on the surface of the GUV, so that the projection on the bead displacement in the x and y axes changed rapidly. Furthermore, because the beads were pre-endocytosed into the GUVs, the initial position of the bead without force was unknown, as compared with the outward-tube pulling assay in which the bead position was recorded before it became attached to the GUV. Therefore, ΔF instead of F was plotted as it is more reliable. In Extended Data Fig. 8d, 6 μM Mgm1 was added to increase protein polymerization and therefore the force generated. In experiments requiring GTP, the buffer was supplemented with 2 mM GTP. The following settings were applied for Fig. 4d, Extended Data Fig. 8a, b: resolution: 512 × 512 × 10 s, 145 nm per pixel, 16 bit; fluorochromes: fluorescein (excitation: 488 nm, bandpass filter 520/50, dichroic beamsplitter 405/488/568/647; LUT: green (Fiji)) rhodamine B (excitation: 561 nm, bandpass filter 607/30, dichroic beamsplitter 405/488/568/647; LUT: red (Fiji)); experiments were performed at room temperature in 20 mM HEPES/NaOH pH 7.4, 200 mM NaCl and 1 mM $MgCl_2$. For Fig. 5c, Extended Data Fig. 8c, d: resolution: 512 × 512 × 30 s, 145 nm per pixel, 16 bit; fluorochromes and experimental conditions as above.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

The atomic coordinates of Mgm1 have been deposited in the Protein Data Bank with accession number 6QL4. Maps obtained by subtomogram averaging were deposited in the Electron Microscopy Data Bank with accession numbers EMD-10062 (with PDB accession number 6RZT for the molecular model) and EMD-4584 for nucleotide-free Mgm1 on the outside of lipid tubes in a close-up view, and the overall tube structure, respectively. EMD-10063 (with PDB 6RZU) shows Mgm1 on the outside of a lipid tube in the GTPγS bound state. EMD-10064 (with PDB 6RZV) and EMD-10065 (with PDB 6RZW) show Mgm1 decorating the inside of a tube without and with GTPγS, respectively. All source data associated with the paper (beyond those deposited) are provided as Supplementary Information.

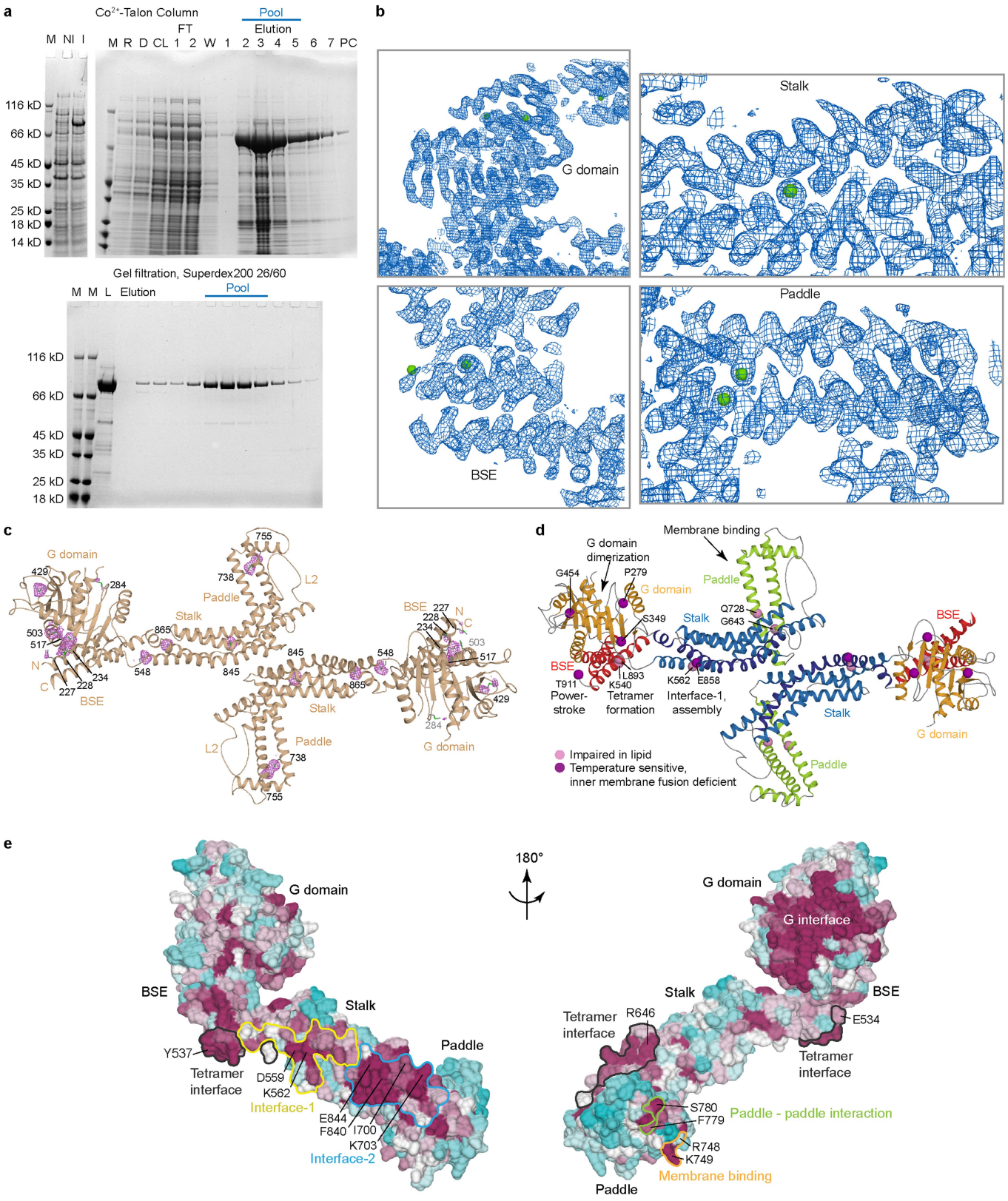
60. Meglei, G. & McQuibban, G. A. The dynamin-related protein Mgm1p assembles into oligomers and hydrolyzes GTP to function in mitochondrial membrane fusion. *Biochemistry* **48**, 1774–1784 (2009).
61. Roux, A. et al. Membrane curvature controls dynamin polymerization. *Proc. Natl. Acad. Sci. USA* **107**, 4141–4146 (2010).
62. Rujiviphat, J. et al. Mitochondrial genome maintenance 1 (Mgm1) protein alters membrane topology and promotes local membrane bending. *J. Mol. Biol.* **427**, 2599–2609 (2015).
63. Mühleip, A. W. et al. Helical arrays of U-shaped ATP synthase dimers form tubular cristae in ciliate mitochondria. *Proc. Natl Acad. Sci. USA* **113**, 8442–8447 (2016).
64. Tarasenko, D. et al. The MICOS component Mic60 displays a conserved membrane-bending activity that is necessary for normal cristae morphology. *J. Cell Biol.* **216**, 889–899 (2017).
65. Barbot, M. et al. Mic10 oligomerizes to bend mitochondrial inner membranes at cristae junctions. *Cell Metab.* **21**, 756–763 (2015).
66. Bohnert, M. et al. Central role of Mic10 in the mitochondrial contact site and cristae organizing system. *Cell Metab.* **21**, 747–755 (2015).
67. Hessenberger, M. et al. Regulated membrane remodeling by Mic60 controls formation of mitochondrial crista junctions. *Nat. Commun.* **8**, 15258 (2017).
68. Lee, H., Smith, S. B. & Yoon, Y. The short variant of the mitochondrial dynamin OPA1 maintains mitochondrial energetics and cristae structure. *J. Biol. Chem.* **292**, 7115–7130 (2017).

Acknowledgements This project was supported by ERC grants MitoShape (ERC-2013-CoG-616024 to O.D.) and ScaleCell (ERC- CoG-772230 to F.N.), grants from the Deutsche Forschungsgemeinschaft (SFB958/A12 and SFB740/C07 to O.D., SFB958/A04 and SFB740/D07 to F.N., SFB894/A20 to M.v.d.L., IRTG1830 to M.v.d.L. and F.W., SFB807 to R.S.), the Max Planck Society, a Humboldt fellowship to J.K.N., a pre-doctoral fellowship of the Boehringer Ingelheim Fonds to F.W., a Sofja Kovalevskaja Award from the Alexander von Humboldt Foundation to M.K., and a DOC Fellowship of the Austrian Academy of Sciences to M.H. We thank Y. Roske for help with crystallographic data collection, structure solution and Isothermal titration calorimetry measurements, T. Brandt for help and assistance in preparing cryo-EM samples, D. Mills for cryo-EM maintenance, B. Purfürst for support in the negative-stain EM analyses, T. Bock-Bierbaum for helpful comments on the manuscript, E. Werner from Research Network Services for his careful work on the videos, A. Xavier for help with Mgm1 fluorescence labelling, and the entire BESSY team for generous support during data collection at beamlines BL14.1, BL14.2 or BL14.3.

Author contributions K.F. designed the construct, grew the crystals and solved the structure. L.D. determined the cryo-ET reconstructions with support from A.M., R.S. and M.K.; J.K.N. and F.N. conducted and analysed molecular modelling and molecular dynamics simulations; F.W. and A.v.d.M. performed yeast-growth assays; and A.-K. P. together with N.C. carried out the tube-pulling assay. J.K.N., F.W. and A.-K.P. contributed equally to this study. J.S. purified the protein and J.S. and K.F. carried out the liposome co-sedimentation and GTPase assays; H.L. performed the analytical ultracentrifugation assays; E.R. and M.H. grew initial crystals of related Mgm1 constructs; C.M. and S.K. analysed yeast mitochondria using electron microscopy; K.F., L.D., J.K.N., C.M., A.R., M.v.d.L., W.K. and O.D. designed the research and interpreted structural data. K.F., L.D., J.K.N., M.v.d.L., W.K. and O.D. wrote the manuscript.

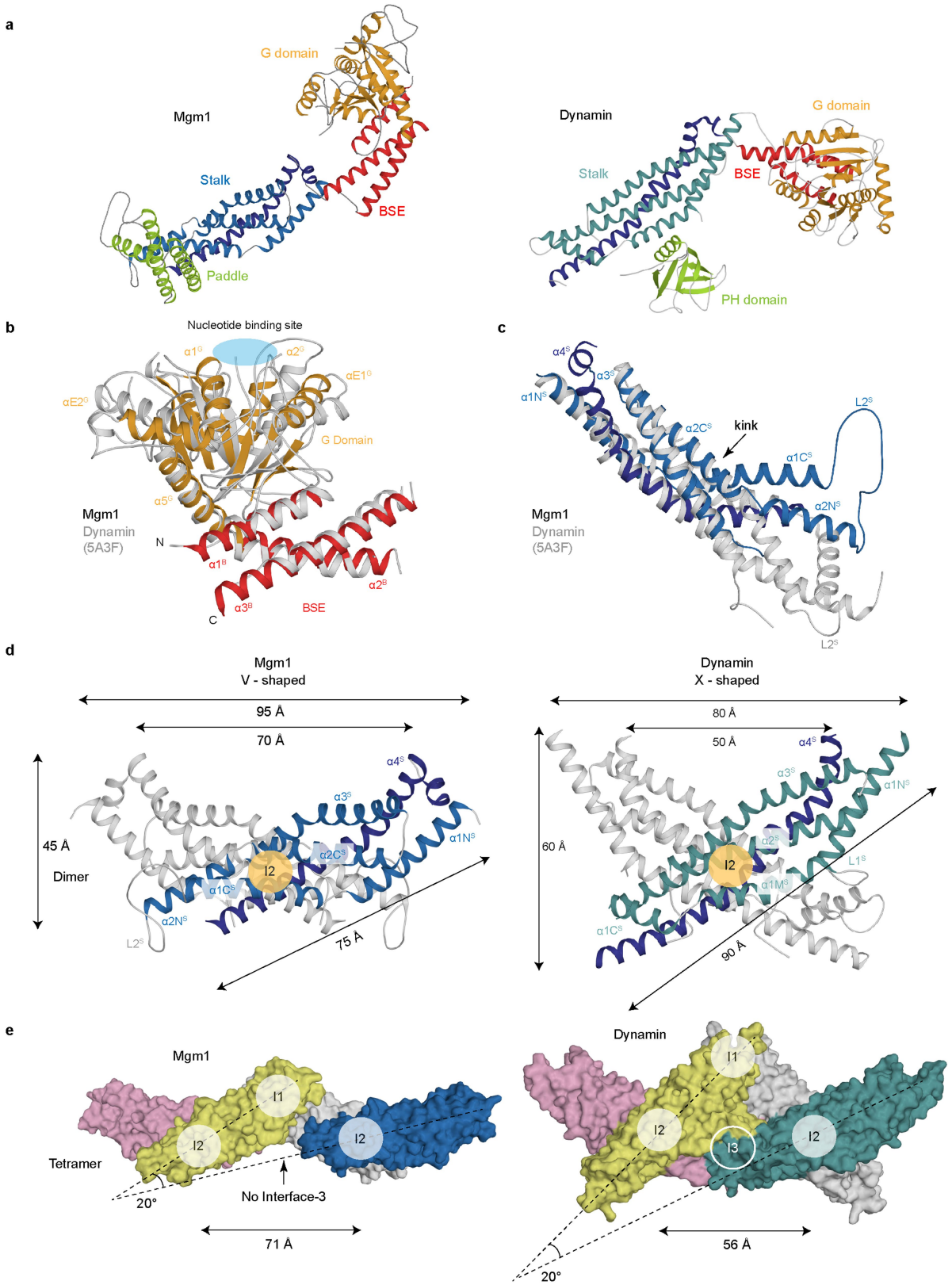
Competing interests The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41586-019-1372-3.
Correspondence and requests for materials should be addressed to K.F. or W.K. or O.D.
Peer review information *Nature* thanks Harry Low, Tom Shemesh and the other anonymous reviewer(s) for their contribution to the peer review of this work.
Reprints and permissions information is available at http://www.nature.com/reprints.

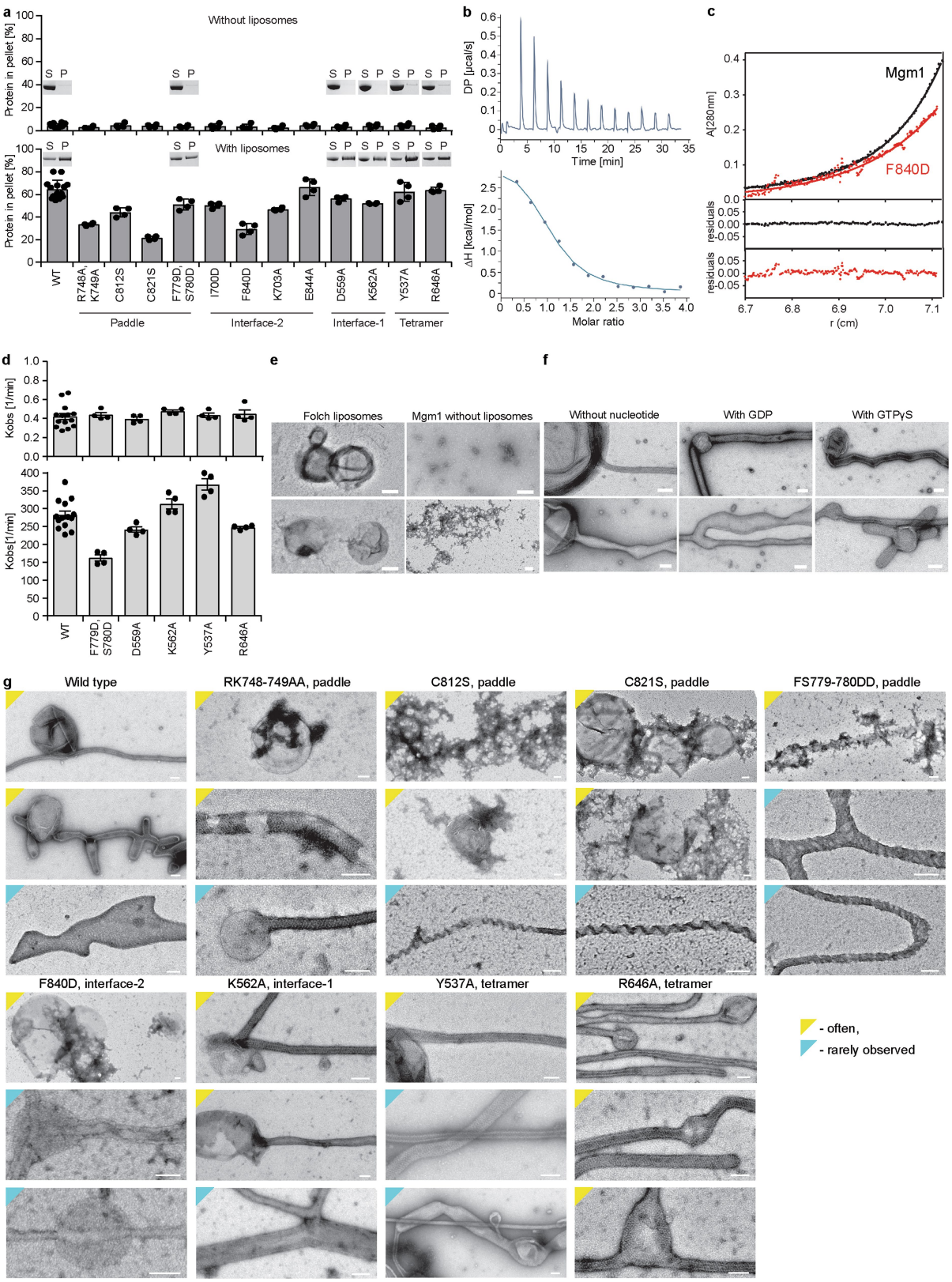


Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Structure determination and analysis. **a**, SDS–PAGE of recombinantly expressed and purified Mgm1. M, marker proteins; NI, whole-cell lysate, non-induced; I, whole-cell lysate, induced; R, whole-cell lysate, resuspended, collected cells; D, whole-cell lysate, disrupted cells; CL, cleared lysate; FT, flow-through; W, buffer wash; PC, after cleavage by PreScission Protease; L, as loaded onto gel filtration column ($n = 5$ independent experiments). **b**, Selenium sites and experimental density at 1.4σ before model building and refinement of the G domain (top left), stalk (top right), BSE (bottom left) and paddle domain (bottom right). **c**, Ribbon diagram of Mgm1 dimer, indicating the positions of confirmed methionines in ball-and-stick representation. Anomalous difference density is contoured at 2.5σ in magenta. An anomalous difference map was calculated from refined phases, resulting in discrete difference peaks indicating the positions of selenium atoms. Four selenium sites in the G domain, three in the BSE, two in the paddle domain and three in the stalk were used to determine the structure and verify the sequence assignment in the model. **d**, Mutations resulting in impaired lipid binding[60] or in temperature-sensitive inner mitochondrial membrane fusion deficits[10] were mapped onto the crystal structure. Mutations localize to the G interface, the G domain/BSE interface, stalk interface-1 or the paddle domain. **e**, Sequence conservation of nine Mgm1 sequences (see Supplementary Fig. 1 for alignments) was plotted on the surface of an Mgm1 monomer. Magenta, high conservation; cyan, low conservation. Residues investigated in this study are labelled and interfaces and contact sites are circled.

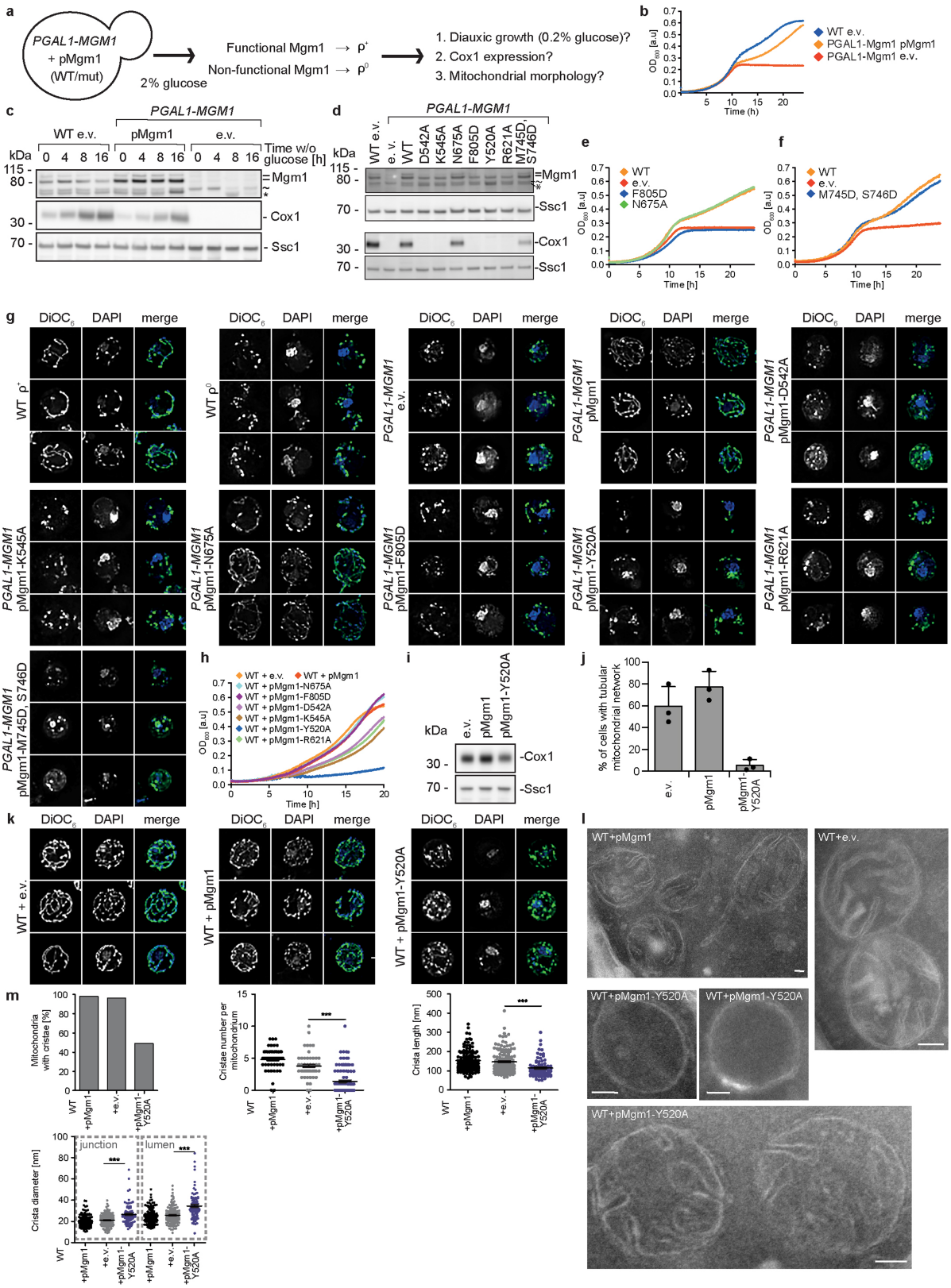


Extended Data Fig. 2 | Comparison of Mgm1 and dynamin. **a**, Monomers of Mgm1 (left) and dynamin (right) coloured by domain. **b**, The G domain and BSE domain of nucleotide-free Mgm1 and dynamin (grey, PDB: 5A3F) were superimposed on the BSE domains with a C_α root-mean-square deviation (r.m.s.d.) of 2.6 Å and 40% sequence identity. Both structures are in the closed state. The nucleotide-binding site is indicated. **c**, Superposition of the upper part of the stalk between Mgm1 and dynamin. In contrast to dynamin, the stalk in Mgm1 is kinked. **d**, Comparison between the stalk dimers of Mgm1 (left) and dynamin (right). In both proteins, the dimer buries a total surface area of 1,200 $Å^2$. However, in Mgm1, interface-2 is shifted towards the paddle, resulting in a V-shaped dimer, whereas the dynamin dimer is X-shaped. **e**, Association of two dimers in the respective tetrameric crystal structures. In dynamin, the assembly of dimers occurs via two interfaces (interface-1 and interface-3), whereas only interface-1 is present in Mgm1.



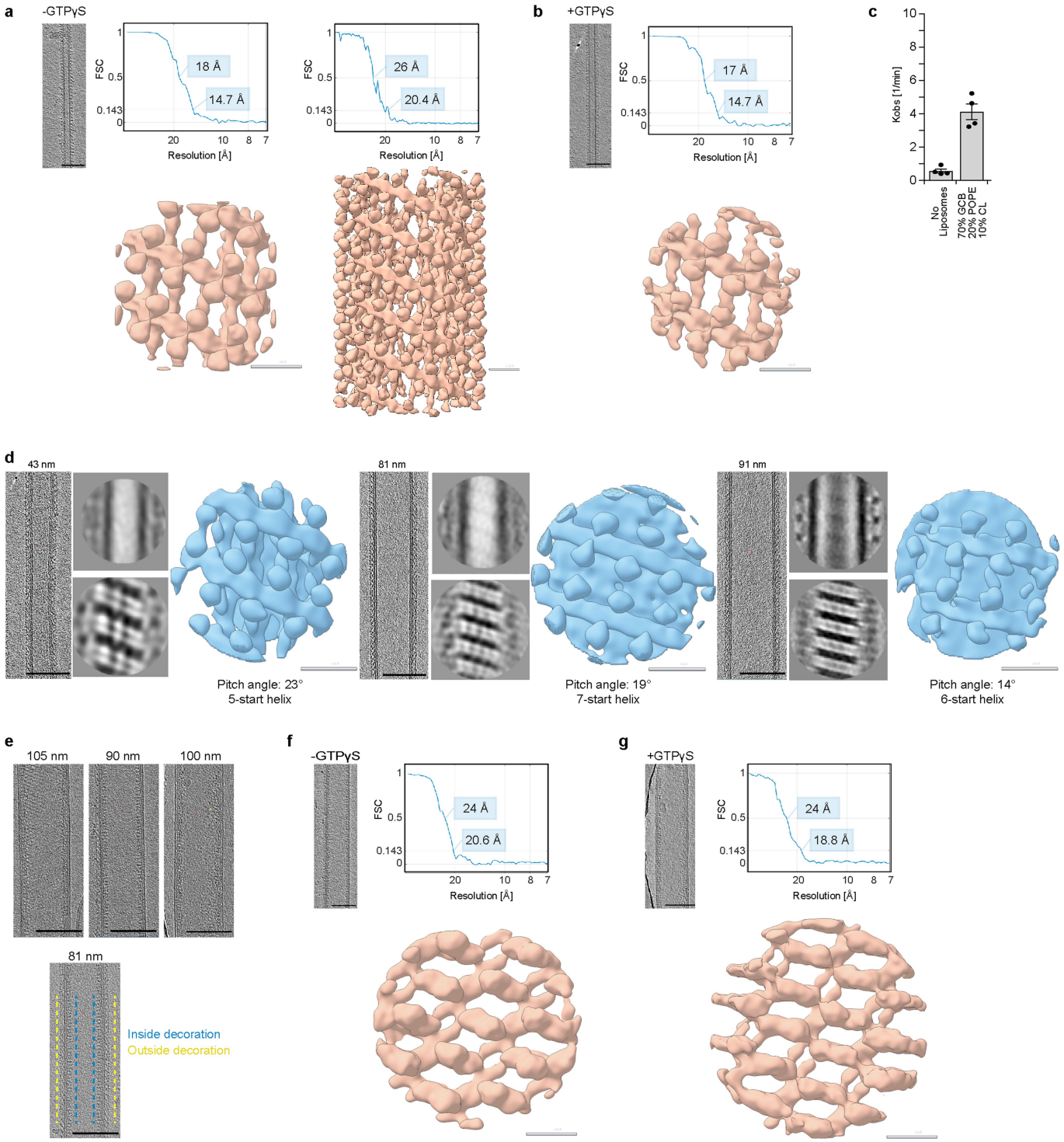
Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Biochemical and negative-stain electron microscopy analysis. a, Liposome-binding assays (see also Figs. 1c, 2d) and quantification for Mgm1 mutants. Error bars indicate s.d. of 4 independent measurements. **b**, Isothermal titration calorimetry experiments showing binding of GTPγS to Mgm1 with a K_d of 9 ± 3 μM, binding number $n = 1.01$, deviation represents root-mean-square (r.m.s.) error of the fit ($n = 1$). **c**, Sedimentation equilibrium of wild-type Mgm1 (black) and Mgm1(F840D) (red) was performed at a protein concentration of 1 mg ml^{-1} at 8,000 r.p.m. and 20 °C. The protein distribution in the cell was monitored by absorbance at 280 nm. Solid lines represent fits to a molecular mass of $M_r = 146 \pm 6$ kDa for wild-type Mgm1 and 78 ± 5 kDa for the Mgm1(F840D) (deviation represents r.m.s. error of the fit), indicating dimeric and monomeric association states at given conditions. The upper panel shows the original data and fits, the lower panels show the residuals from fit to data. **d**, GTPase assays using HPLC analysis. Error bars show s.d. of the mean of 4 independent experiments (each with 4 or 5 data points). **e**, Control experiments for negative-stain electron microscopy analysis of Mgm1-mediated membrane remodelling. Scale bars, 200 nm. **f**, Mgm1 binds to liposomes and forms tubes of different diameters with or without nucleotides present. Scale bars, 100 nm. **g**, Representative electron micrographs for Mgm1 mutants. Proteins with mutations in dimer interface-2 (F840D), in the membrane-binding site (R748A/K749A), the disulfide bond in the paddle domain (C812S and C821S) or in the putative paddle–paddle contact (F779D/S780D) show severe defects in tube formation or in the assembly of a regular liposome decoration compared to Mgm1. Scale bars, 100 nm. $n = 2$ independent experiments for **e**–**g**.

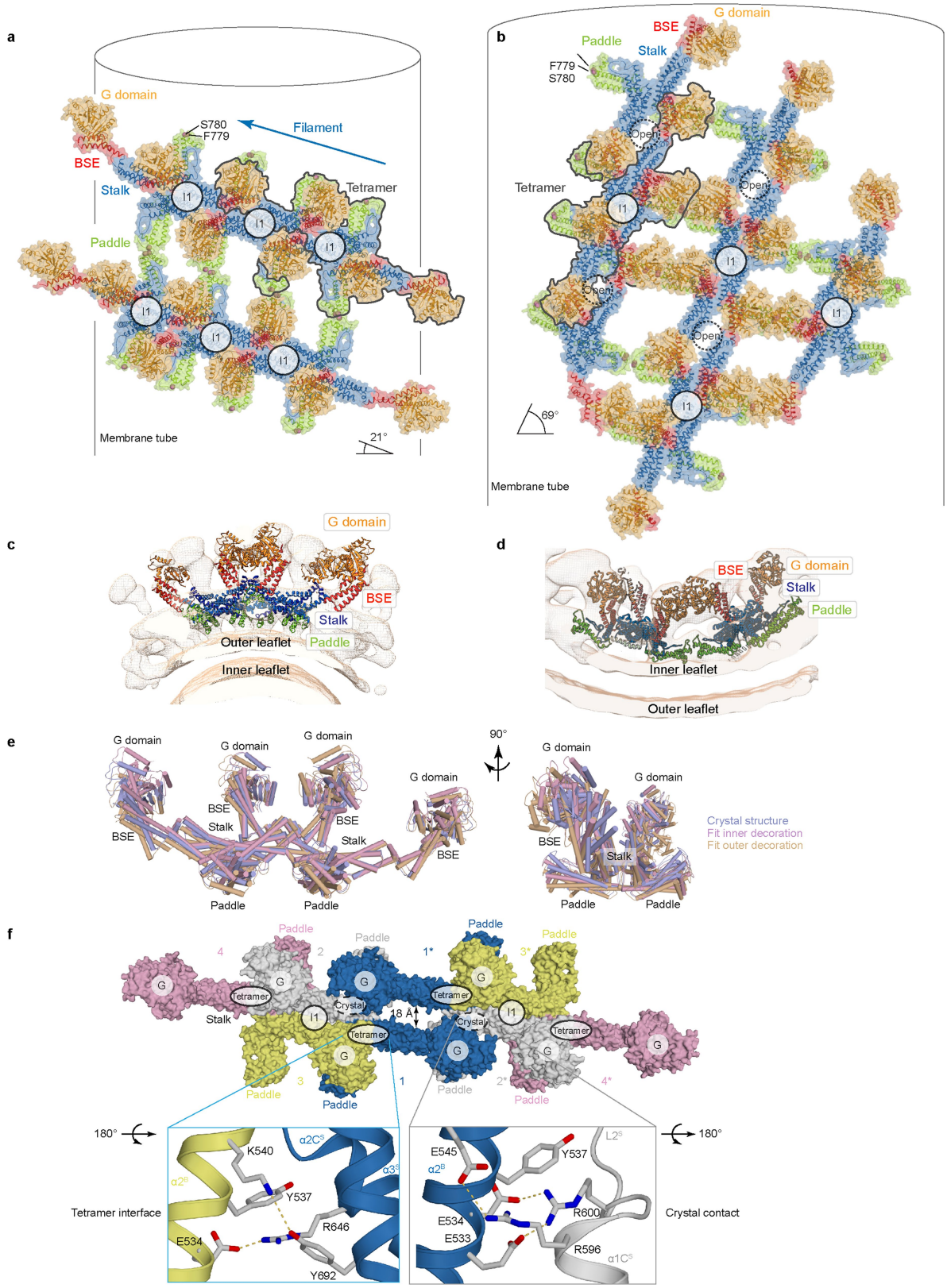


Extended Data Fig. 4 | See next page for caption.

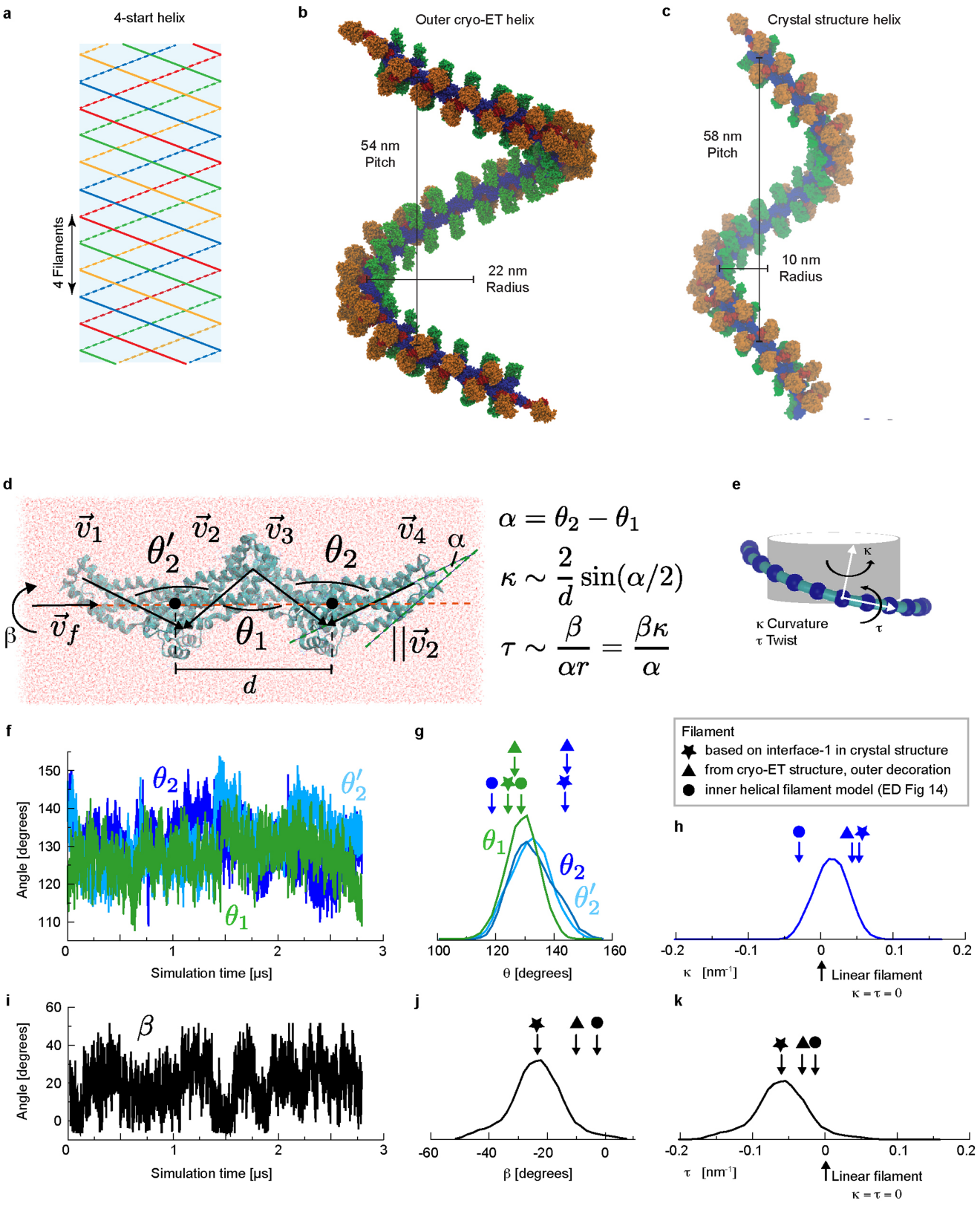
Extended Data Fig. 4 | Yeast assays. a, Schematic overview of yeast complementation experiments. In the presence of 2% glucose, expression of chromosomally encoded Mgm1 from the *GAL1* promoter is suppressed. Yeast cells irreversibly lose the mitochondrial genome in the absence of Mgm1 (that is, become ρ^0) and cannot switch to respiratory growth upon glucose depletion (as shown by the shift from low glucose conditions to the oxidation of ethanol produced during the fermentation of glucose). By co-expressing wild-type yeast Mgm1 or the corresponding Mgm1 mutants, functionality of the Mgm1 variants is assessed through various rescue parameters. **b**, Representative growth curve for the unmodified yeast strain transformed with an empty vector (e.v.), the engineered yeast strain (P_{GAL1}-*MGM1*) complemented with yeast Mgm1 or an empty vector control ($n = 3$ independent experiments). **c**, Time-dependent expression of Mgm1, mitochondrially encoded cytochrome *c* oxidase subunit 1 (Cox1) and the nuclear-encoded mitochondrial heat shock protein Ssc1 (loading control) was assessed by western-blot analysis of isolated mitochondria upon transfer of yeast cells from a glucose-rich to a glucose-depleted medium containing 2% ethanol as the carbon source. = marks the long and short isoforms of Mgm1, ∼ is an unspecific band and * marks an Mgm1 degradation product ($n = 2$ independent experiments). Uncropped blots are shown in Supplementary Fig. 2. **d**, Western-blot analysis of isolated mitochondria from P_{GAL1}-*MGM1* yeast grown in glucose-containing medium containing plasmids that encode the respective mutant ($n = 3$ independent experiments). **e**, **f**, Yeast growth complementation assays with Mgm1 mutants containing mutations in the dimer interface and the paddle–paddle contacts. F805D in yeast corresponds to F840D in *C. thermophilum* and N675A corresponds to I700D. F779D/S780D in *C. thermophilum* corresponds to M745D/S746D in yeast. Representative growth curves are shown from $n = 3$ independent experiments. Data in Fig. 3b and Extended Data Fig. 4e are derived from the same experiment; the controls are shown in all graphs as a reference. **g**, Mitochondrial morphology of the indicated yeast strains was assessed by fluorescence microscopy. DNA and mitochondria were stained with DAPI and $DiOC_6$, respectively. Three representative images from $n = 2$ independent cultures are shown. Dimensions of the images are 7.5 μm × 7.5 μm. **h**, Overexpression of Mgm1(Y520A) (with a mutation in the tetramer interface) leads to a strong dominant-negative effect on respiratory yeast growth (in media containing 3% glycerol as the carbon source). Representative growth curves are shown from $n = 3$ independent experiments. **i**, Overexpression of Mgm1(Y520A) leads to only a partial loss of mitochondrial DNA, as assayed by Cox1 expression. $n = 3$ independent experiments. **j**, **k**, Overexpression of dominant-negative Mgm1(Y520A) leads to a fragmentation of the mitochondrial network. Representative images and quantification of mitochondrial morphology in cells from $n = 3$ independent cultures, data displayed as mean ± s.d. **l**, Representative electron micrographs of yeast ultrathin sections assaying mitochondrial ultrastructure. Compared to mitochondria in wild-type yeast transformed with empty vector or *pMgm1*, mitochondria from cells expressing Mgm1(Y520A) showed a substantial loss of cristae and altered crista shape, as indicated by an increased diameter of the crista junctions and lumen and shorter crista length. Scale bars, 70 nm. **m**, Quantification of cristae morphology. WT+pMgm1: $n_{mito} = 208$, $n_{cristae} = 132$; WT+e.v.: $n_{mito} = 201$, $n_{cristae} = 135$; WT+pMgm1(Y520A): $n_{mito} = 202$, $n_{cristae} = 81$; 2 independent experiments. $^{***}P < 0.0001$ (Gaussian approximation); Mann–Whitney *U*-test (two-sided, 95% confidence interval); cristae number graph shows mean ± s.e.m.: WT+pMgm1: (4.8 ± 0.2) nm; WT+e.v.: (3.8 ± 0.2) nm; WT+pMgm1(Y520A): (1.4 ± 0.2) nm; crista length graph shows mean ± s.e.m.: WT+pMgm1: (153 ± 5) nm; WT+e.v.: (147 ± 5) nm; WT+pMgm1(Y520A): (115 ± 5) nm; crista diameter graph shows mean ± s.e.m.: WT+pMgm1 junction: (19.9 ± 0.5) nm; WT+e.v. junction: (21.0 ± 0.5) nm; WT+pMgm1(Y520A) junction: (26 ± 1) nm; WT+pMgm1 lumen: (24.7 ± 0.6) nm; WT+e.v. lumen: (25.8 ± 0.7) nm; WT+pMgm1(Y520A) lumen: (35 ± 2) nm.



Extended Data Fig. 5 | Cryo-ET analysis. a, **b**, **f**, **g**, Electron micrographs on the left show one tomographic slice of each sample. The density maps below obtained by subtomogram averaging are bandpass-filtered to the Fourier pixel value at 0.143 of the FSC curve. The masked FSC curves of each subtomogram average are indicated with resolutions obtained at 0.5 and 0.143 FSC. **a**, Mgm1 on the outside of a galactocerebroside-containing lipid tube in the apo form. On the right, a larger box size was used for processing in order to visualize the complete protein coat decorating the lipid tube. **b**, Mgm1 in the GTPγS-bound form on the outside of galactocerebroside-containing lipid tubes are very similar to the apo form, whereas nucleotide-free dynamin assembles differently compared to the guanosine-5′-[(β,γ)-methyleno]triphosphate-bound form[22]. **c**, GTPase assays of Mgm1 in the presence of lipid tubes containing galactocerebroside, $n = 4$, errors represent s.d. from the mean. **d**, Low-resolution cryo-ET reconstructions of GTPγS-bound Mgm1 assembled on the outside of Folch membrane tubes of different diameters, as measured between bilayer centres. On the basis of the pitch angle θ and the tube diameter d, the number of helical repeats (n-start) was estimated as $n = 2\pi r\tan\theta/h$, where the filament radius $r = d/2+4$ nm and the width from paddle tip to tip h is 13 nm. Although the basic filament architecture appears very similar, the filaments adapt their orientation to the curvature of the membrane tube. **e**, Representative electron micrographs showing Mgm1 coating the inner surface of a membrane tube (top) or both sides of the membrane tube (below). **f**, **g**, Cryo-ET reconstruction of Mgm1 in the apo and GTPγS-bound form on the inside of tubulated Folch liposomes, as in **a** and **b**. Grey scale bars, 10 nm; black scale bars, 100 nm.

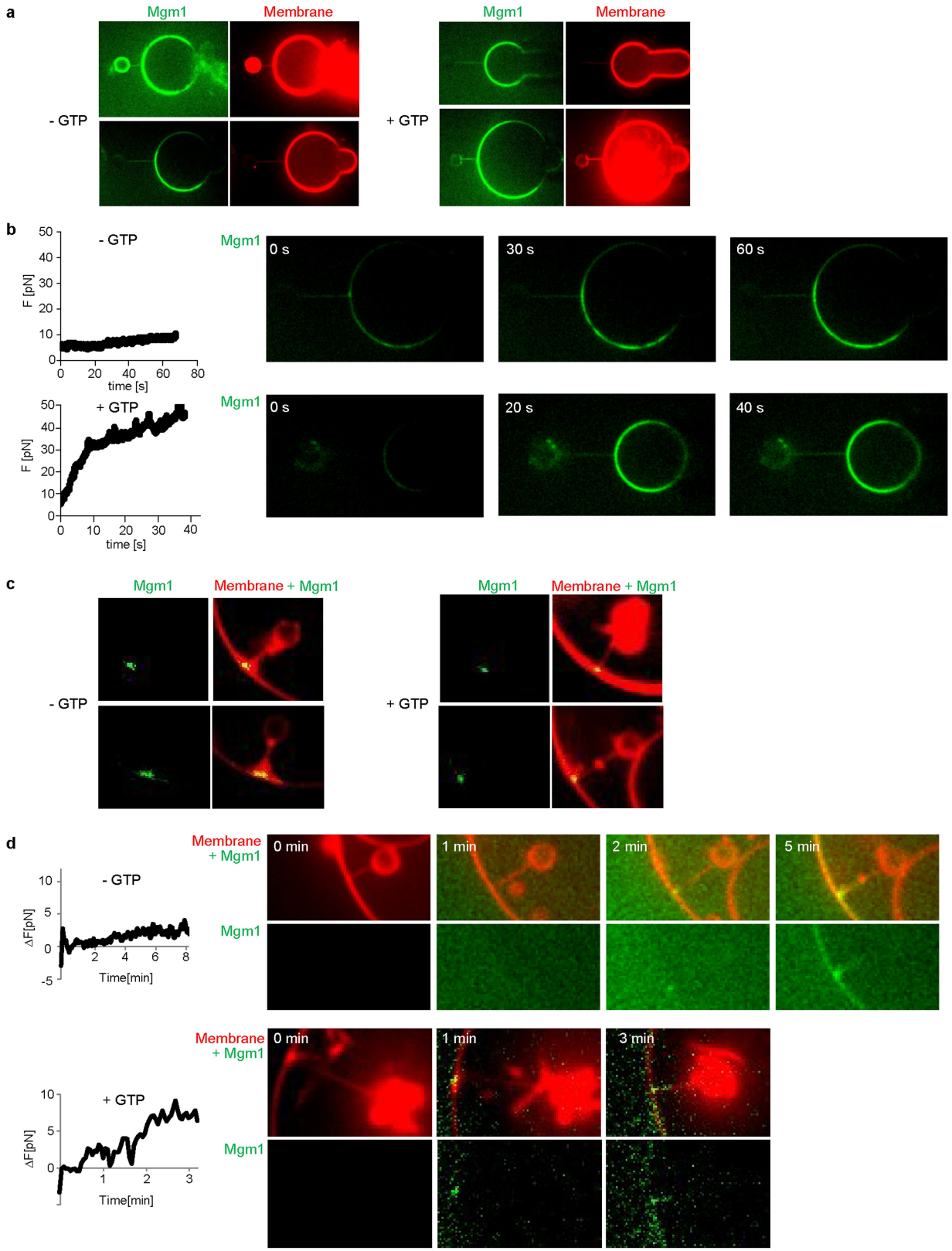


Extended Data Fig. 6 | Mgm1 tetramers in crystal and membrane lattices. a–d, Mgm1 assemblies in the presence of GTPγS on the outer (**a**, **c**) and inner surface (**b**, **d**) of a membrane tube. **a**, **b**, Surface representations of flexibly fitted Mgm1 molecules, showing their arrangement in the protein lattice. **c**, **d**, Fit into the corresponding cryo-ET volume. Note that the membrane density and, consequently, the paddle–membrane contact, is more prominent in the GTPγS-bound form compared with the nucleotide-free form (Figs. 4b, 5a). **e**, Comparison of Mgm1 tetramers in the crystal lattice (blue) with tetramers fitted to the subtomogram average volumes obtained for the external (orange) and internal surface lattice (pink). Fitting the paddle and the BSE and G domains required only minor rearrangements. **f**, Tetramers in the crystal lattice pack into a linear assembly. Crystal contacts between two tetramers are mediated by the BSE domain of one tetramer (blue) and the stalk domain of the neighbouring tetramer (grey), resulting in an open interface-1. When comparing intra- and inter-tetramer interactions, BSE domain residues E533, E534 and Y537 in $\alpha2^B$ bind to different sites of the adjacent stalks.



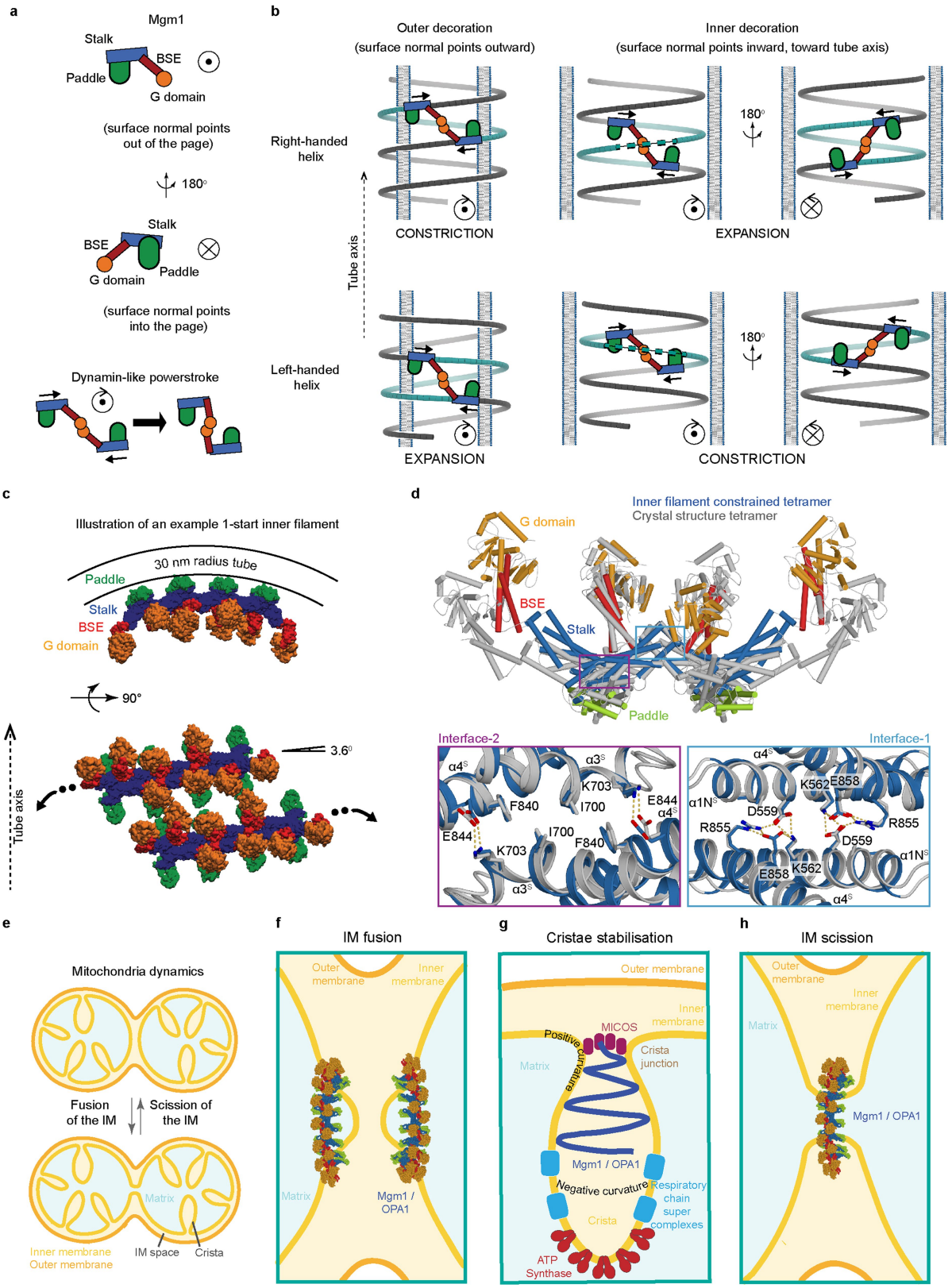
Extended Data Fig. 7 | See next page for caption.

Extended Data Fig. 7 | Molecular dynamics simulations. a, Schematic of a 4-start helix. **b**, Mgm1 filaments in a 4-start helix, as in the cryo-ET volume on the outside of lipid tubes. The filament is defined as a continuous string of stalk domains connected by alternating interface-1 and interface-2. With this arrangement, filaments have a radius of 22 nm (axis to the centre of the stalk) and pitch of 54 nm. **c**, A string of dimers in contact through identical interfaces-1, as in the crystal structure, results in a left-handed helical arrangement with a large pitch, similar to the cryo-ET filament of the outside decoration. **d**, Snapshot of the stalk tetramer structure in the molecular dynamics simulation box. Analysis of the stalk tetramer conformation in molecular dynamics simulations gives information about the structural preferences of the filament in the absence of other domains. Geometrical parameters are drawn on the structure. d is the distance between the centres of mass of neighbouring dimers (marked as filled black circles). 95% of the variation in d is between 6.8 and 7.7 nm. $\boldsymbol{v}_1$–$\boldsymbol{v}_4$ are vectors pointing along each stalk monomer, defining angles θ_1, θ_2, and θ_2' as shown. α is the net in-plane rotation defined by $\boldsymbol{v_2} \times \boldsymbol{v}_3$, and is related to the local radius of curvature of a filament containing the tetramer. α can be simply written as a difference of the two interface angles, $\alpha = \theta_2 - \theta_1$, where positive/negative α implies positive/negative curvature; $\theta_2 > \theta_1$ results in positive curvature and $\theta_2 < \theta_1$ results in negative curvature. β is the relative rotation angle of one dimer relative to the next, which controls the pitch and, therefore, the handedness of the helix. β is defined by the angle between the vectors $\boldsymbol{v}_1 \times \boldsymbol{v}_2$ and $\boldsymbol{v}_3 \times \boldsymbol{v}_4$ viewed along $\boldsymbol{v}_\mathrm{f}$. $\boldsymbol{v}_\mathrm{f}$ is a unit vector in the direction of the filament defined by connecting the centres of mass of the two dimers. The elastic coordinates of a helical filament are the curvature κ and the twist τ. Positive/negative κ yields helices that bind to positive/negative membrane curvature. κ and τ can be approximately related to α and β, and the relations are indicated in the figure. **e**, Schematic of the curvature κ and the twist τ. For helices with a low pitch, κ is approximately the inverse radius of curvature ($1/r$). **f**, The angles θ_1, θ_2 and θ_2' are plotted over a portion (2.8 μs out of a total of 12 μs) of the simulation period. **g**, Distributions of θ_1, θ_2 and θ_2' over the whole simulation period. θ_2 and θ_2' are, in principle, identical and the similarity of the distributions indicates sufficient sampling. In the crystal structure, $\theta_1 = 123°$ and $\theta_2/\theta_2' = 142°/144°$. The flexibilities of interface-1 and interface-2 are similar, as seen from the similar distribution widths. The peak of the θ_1 distribution is centred on the parameters obtained for the crystal packing, whereas θ_2/θ_2' is different, which may indicate that additional domain contacts present in the crystal stabilize a different configuration of interface-2. **h**, Using the relations shown in **d**, θ_1 and θ_2 at each snapshot are used to estimate the distribution of the curvature. The curvature distribution is centred near 0, which indicates that the stalk filament (at zero twist) prefers weakly curved or flat membranes. **i**, The angle β is plotted over a portion (2.8 μs out of a total of 12 μs) of the simulation period. **j**, **k**, The distributions of β (**j**) and τ (**k**) over the whole simulation period. A negative β or τ indicates that the stalk filament prefers a left-handed twist, but right-handed twists are thermally accessible. Note that no substantial correlation is seen between θ_1, θ_2/θ_2' and β.



Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Mgm1 attachment to membranes of different curvature. Tube-pulling experiments, as described in Figs. 4d, 5c. Mgm1 was labelled with a fluorescein tag (green) and GUVs with Rhod-PE (red). Positive force is defined as pointing from the bead to the GUV. **a**, Tubes were pulled outward of single GUVs held by a micropipette ($n = 8$ independent experiments in the absence of GTP, $n = 10$ independent experiments in the presence of GTP). **b**, Representative time-lapse images of nucleation and growth of Mgm1 polymers on tubes pulled away from a GUV (right), and corresponding force measurements (left). **c**, Representative examples for tubes pulled into single GUVs adhering to the glass surface ($n = 7$ independent experiments in the absence of GTP, $n = 7$ independent experiments in the presence of GTP). **d**, Same as in **b**, but for tubes pulled into GUVs. ΔF is shown, as absolute forces were difficult to measure. Although Mgm1 covered the GUV surface in the experiments shown in **c** and **d**, it apparently did not oligomerize along the entire inward-pulled tube, as judged from the fluorescence signal. This probably reflects decreased diffusion of Mgm1 along the tube lumen. However, when the tube is not fully covered, a GTP-dependent shape change of the Mgm1 coat in the tube would not induce a force, as previously demonstrated for dynamin[61]. Therefore, the force increase probably results from the GTP-dependent remodelling of the Mgm1 coat on the GUV. In the case of outer decoration, Mgm1 oligomerizes on the tube and the GUV. In this case, the force increase can be caused by GTP-dependent alterations of the Mgm1 coat on the GUV and/or tube expansion. We note that these experiments gave no hint of GTP-driven constriction of membrane tubes.



Extended Data Fig. 9 | See next page for caption.

Extended Data Fig. 9 | Model of Mgm1 action. a, On the basis of the close similarity of the G domains and BSE domains of Mgm1 and dynamin (Extended Data Fig. 2b), we propose that Mgm1 and dynamin perform similar power strokes. Dimerization of the G domain would link neighbouring Mgm1 filaments. The power stroke would then result in negative torque in the direction of the membrane normal. In **b**, a circle with a dot indicates a vector towards the viewer and a circle with an x indicates vector in the opposite direction. The arrow represents the direction of the torque. Note that power-stroke torque is independent of membrane curvature and helix handedness. During the power stroke, the helix pitch remains constant because of the G domain contacts. Unwinding or winding of filaments then translates into a change in helix diameter. Inter-paddle contacts must be weak or absent as the filaments slide past each other. **b**, The power-stroke torque applies an equal and opposite force between neighbouring turns. For outside decoration, the surface normal points outward. The resulting forces would constrict a right-handed helix and expand a left-handed helix. For inside decoration, the surface normal points inward, reversing the sign of the power-stroke torque. This reverses the resultant forces on the filament, which would expand a right-handed helix and constrict a left-handed helix. See also Supplementary Video 1. **c**, Modelling an example helical Mgm1 filament on an inner-tube surface. Although the Mgm1 tetramer on the inside lattice observed by cryo-ET resembled the crystal tetramer closely, formation of a continuous filament on the inside of a narrow tube would require curvature changes in the tetramer relative to the crystal structure. Using an all-atom structure-based model, we explore how the tetramer structure might change as part of a tight filament. The modelling parameters ensured that a short filament (4 dimers) fits within the steric constraints of a 30-nm-radius tube, and that the pitch results in a 1-start helix (left-handed pitch angle of 3.6°). Otherwise, the shape of the tetramer is free to find its optimal shape. Changes in the interface bending angles result in a transition from positive curvature ($\theta_2 > \theta_1$) to negative curvature ($\theta_2 < \theta_1$) (Extended Data Fig. 7d). **d**, Comparison of the constrained tetramer shown in **c** (central dimers) with the crystal structure. Minor changes in interface-1 and larger changes in interface-2 (with minimal changes to atomic packing, see insets) enable a conformational switch within the tetramer from binding to a concave surface (as in the crystal packing geometry) to binding to a convex surface. In this case, $\theta_1 = 128°$ and $\theta_2 = 117°$. See also Extended Data Fig. 7g for comparison to explicit solvent simulations. **e**, Schematic overview of mitochondrial inner membrane remodelling. **f**–**h**, Models of mitochondrial membrane remodelling by Mgm1 and OPA1 filaments. **f**, During inner-membrane fusion, Mgm1 or OPA1 filaments may assemble on opposing membrane buds to stabilize the membrane curvature at the fusion site, as previously proposed[62]. **g**, On the inner surface of cristae, Mgm1 or OPA1 filaments may assemble into left-handed helical filaments to constrict the crista junction in a GTPase-dependent fashion. Alternatively, they may assemble into right-handed helical filaments that expand the crista volume to prevent their collapse. In this way, Mgm1 filaments may counteract the membrane-constricting activity of the ATPase synthase dimers[63] or the MICOS complex[64–67] to pull lipids into cristae and enable the dynamic transition from a tight crista state with reduced oxidative phosphorylation to an expanded active state with high oxidative phosphorylation activity. In agreement with this model, cristae have been shown to collapse when a GTPase-deficient OPA1 variant is expressed[14]. **h**, Similar to dynamin assemblies at the neck of clathrin-coated pits, Mgm1 or OPA1 may assemble in a right-handed helix around the neck of an inner membrane junction, resulting in constriction and membrane scission upon GTP hydrolysis. The assembly geometry of the Mgm1 or OPA1 filaments may depend on lipid composition, interaction partners or the specific Mgm1 or OPA1 isoform. Consistent with the latter assumption, inner membrane fusion requires the long form of OPA1, but the short OPA1 isoforms are sufficient for stabilizing crista membranes[68].

Extended Data Table 1 | Data collection and refinement statistics

a, Crystallographic data

	Mgm1, SeMet pdb code 6QL4
Data collection	
Space group	$P4_122$, 1 dimer/ASU
Cell dimensions	
a, *b*, *c* (Å)	147.4, 147.4, 344.7
α, β, γ (°)	90, 90, 90
Wavelength	0.9794 Å
Resolution (Å)*	3.60 (3.60-3.69)
R_{sym}* (%)	22.0 (184)
$I / \sigma I$*	8.0 (1.1)
Completeness (%)*	99.8 (98.4)
Redundancy	6.7
Refinement	
Resolution (Å)	49.1 – 3.6
No. reflections	44,814
R_{work} / R_{free}	24.4 / 25.2
No. atoms	
Protein	10,332
Ligand/ion	2
Water	0
B-factors	
Protein	212 Å^2
Ligand/ion	101 Å^2
Water	-
R.m.s deviations	
Bond lengths (Å)	0.004
Bond angles (°)	0.986

b, Cryo-ET data

	Mgm1+GTPγS Outside (EMDB-10063)	Apo Mgm1 Outside (EMDB-10062)	Apo Mgm1 Outside, overall tube (EMDB-4584)	Mgm1+GTPγS Inside (EMDB-10065)	Apo Mgm1 Inside (EMDB-10064)
Data collection and processing					
Magnification	53,000	53,000	53,000	53,000	53,000
Voltage (kV)	300	300	300	300	300
Electron exposure (e^-/Å^2)	90 – 100	90 – 100	90 – 100	90 – 100	90 – 100
Per tomogram					
Defocus range (μm)	2 - 4	2 - 4	2 - 4	2 - 4	2 - 4
Pixel size (Å)	2.7	2.7	2.7	2.7	2.7
Symmetry imposed	No	No	No	No	No
Initial particle images (no.)	71,884	12,440	2,214	4,751	1,874
Final particle images (no.)	9,471	11,474	1,677	1,820	1,792
Map resolution (Å)	14.7	14.7	20.4	18.8	20.6
FSC threshold	0.143	0.143	0.143	0.143	0.143

a, Data collection and refinement statistics for the crystallographic data.
b, Data collection and processing statistics for cryo-ET data.
*Values in parentheses are for the highest-resolution shell. The data were collected from a single crystal.

nature research

Corresponding author(s): Oliver Daumke

Last updated by author(s): May 24, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	MXCube v1, XDS, SerialEM-3.5 and EM-3.6, Leica Application Suite X 3.4, Tecan SparkControl V2.1
Data analysis	XDS, Phenix-1.11.1-2575, Coot 0.8, Unblur (v1.0.2), MotionCor2 (v1.10-Cuda8.0), IMOD (v4.8.56 and v4.10.9), DYNAMO (v1.1.226), MATLAB (MATLAB R2018b v9.5), relion_reconstruct (Relion v2.1), UCSF Chimera (v1.12), Gromacs 4.5.3, Acemd 2016.3, SMOG2 v2.1, VMD1.9.3, Graphpad Prism 6.0h, Huygens Essential 17.04, Microsoft Excel 2013, Fiji/ImageJ 1.51s, Adobe Illustrator CS 06

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Structure coordinates have been deposited in the Protein Data Bank (6QL4) and cryo-ET maps have been deposited at the Electron Microscopy Data Bank (EMD-4582 to 4586). All source data associated with the paper (beyond those deposited) are provided as supplementary information. Other data or informations are available on request from the corresponding authors.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to predetermine sample size. The determined sample size was adequate as the differences between experimental groups was reproducible, as indicated. X-ray diffraction data were collected until completeness of the data set. For cryoEM, sample sizes were determined by available electron microscopy time and density of particles on the electron microscopy grids. Sample sizes were sufficient to obtain structures at the reported resolution.
Data exclusions	Tomograms showing imaging problems were excluded from the data set. Otherwise, no data were excluded.
Replication	The experimental findings were reproduced in multiple independent experiments. The number of independent experiments and biological replicates in each data panel is indicated in the figure legends. CryoEM data sets were collected in different sessions, on different days from repetitions of the same experimental setup (see methods part). For liposome binding and stimulated GTPase assays, at least two different liposomes and protein batches were used for each reported data point.
Randomization	Randomization was not formerly performed in this study as it did not involve animals and/or human research participants. When generating yeast strains, random single clones were selected for further experiments.
Blinding	Blinding is not relevant for protein structure determination and cryoET map interpretation since the results are not subjective. Investigators were not blinded during functional assays. Importantly, data collection during yeast growth experiments (using a microplate reader) was automated and non-subjective.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used	If not otherwise mentioned, all used antibodies are custom-made polyclonal antisera. Antibodies against Mgm1 (GR796-5), Mic10 (GR3367-7), Cox1 (GR1538-4), Ssc1 (GR119-3), described in Ieva et al., 2014, Malsburg et al. 2011 and Morgenstern et al., 2017, were raised in rabbit using synthetic peptides and diluted 1:500 in TBST-5% (w/v) milk. HRP-conjugated secondary antibodies were obtained from Merck/Millipore (catalogue number AP187P) and diluted 1:5000 in TBST-5% (w/v) milk.
Validation	Antibodies were validated by Western blotting using yeast deletion mutants.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The YPH499 wild-type yeast strain has been described (Sikorski & Hieter, 1989) and was obtained from the laboratory of Nikolaus Pfanner (University of Freiburg/Germany).
Authentication	The PGAL1-MGM1 mutant yeast strain was generated as described in the methods section and validated by control PCR as well as Western blotting after culturing on galactose-containing media.

Mycoplasma contamination	In this study, no cells were used that can be contaminated with Mycoplasma.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in this study.

CORRECTIONS & AMENDMENTS

CORRECTION
https://doi.org/10.1038/s41586-019-1298-9

Author Correction: Forearc carbon sink reduces long–term volatile recycling into the mantle

P. H. Barry, J. M. de Moor, D. Giovannelli, M. Schrenk, D. R. Hummer, T. Lopez, C. A. Pratt, Y. Alpízar Segura, A. Battaglia, P. Beaudry, G. Bini, M. Cascante, G. d'Errico, M. di Carlo, D. Fattorini, K. Fullerton, E. Gazel, G. González, S. A. Halldórsson, K. Iacovino, J. T. Kulongoski, E. Manini, M. Martínez, H. Miller, M. Nakagawa, S. Ono, S. Patwardhan, C. J. Ramírez, F. Regoli, F. Smedile, S. Turner, C. Vetriani, M. Yücel, C. J. Ballentine, T. P. Fischer, D. R. Hilton & K. G. Lloyd

Correction to: *Nature* https://doi.org/10.1038/s41586-019-1131-5, published online 24 April 2019.

In this Article, the original affiliation 2 (California Water Science Center, USGS, San Diego, CA, USA), associated with authors P. H. Barry and J. T. Kulongoski, was not applicable because the study was not directly funded or approved by the United States Geological Survey (USGS). This affiliation has been removed and all subsequent affiliations have been renumbered accordingly. In addition, in the Acknowledgements, the surname of J. Labidi was misspelled 'Libidi', and the following statement should have been present: "J.M.d.M. acknowledges funding from Universidad Nacional Costa Rica, the World Bank, and the Costa Rican Ley Transitorio 8933 used to acquire a laser carbon isotope system in collaboration with R. Sánchez-Murillo and G. Esquivel-Hernandez". The Article has been corrected online.

CORRECTION

https://doi.org/10.1038/s41586-019-1342-9

Author Correction: Climatic controls of decomposition drive the global biogeography of forest–tree symbioses

B. S. Steidinger, T. W. Crowther, J. Liang, M. E. Van Nuland, G. D. A. Werner, P. B. Reich, G. J. Nabuurs, S. de-Miguel, M. Zhou, N. Picard, B. Herault, X. Zhao, C. Zhang, D. Routh, K. G. Peay & GFBI consortium

Correction to: *Nature* https://doi.org/10.1038/s41586-019-1128-0, published online 15 May 2019.

In this Letter, the middle initial of author G. J. Nabuurs was omitted, and he should have been associated with an additional affiliation: 'Forest Ecology and Forest Management Group, Wageningen University and Research, Wageningen, The Netherlands' (now added as affiliation 182).

In addition, the following two statements have been added to the Supplementary Acknowledgements. (1): 'We would particularly like to thank The French NFI for the work of the many field teams and engineers, who have made extraordinary efforts to make forest inventory data publicly available.' (1): 'Sergio de Miguel benefited from a Serra-Húnter Fellowship provided by the Generalitat of Catalonia.'

Finally, the second sentence of the Methods section should have cited the French NFI, which provided a national forestry database used in our analysis, to read as follows: 'The GFBi database consists of individual-based data that we compiled from all the regional and national GFBi forest-inventory datasets, including the French NFI (IGN—French National Forest Inventory, raw data, annual campaigns 2005 and following, https://inventaire-forestier.ign.fr/spip.php?rubrique159, site accessed on 01 January 2015)'. All of these errors have been corrected online.

CORRECTION

https://doi.org/10.1038/s41586-019-1366-1

Author Correction: Sampling the volatile-rich transition zone beneath Bermuda

Sarah E. Mazza, Esteban Gazel, Michael Bizimis, Robert Moucha, Paul Béguelin, Elizabeth A. Johnson, Ryan J. McAleer & Alexander V. Sobolev

Correction to: *Nature* https://doi.org/10.1038/s41586-019-1183-6, published online 15 May 2019.

In this Letter, two of the authors (S.E.M. and E.G.) noticed errors in Extended Data Fig. 5, the legend of Fig. 2, the legend of Extended Data Fig. 7 and reference 41. In the key of Extended Data Fig. 5 the pink (carbonated eclogite experiments) and yellow (carbonated peridotite experiments) star symbols have been swapped. In the legend of Fig. 2c, "and eclogite" has been deleted from the text "with experimental data from melting carbonated peridotite and eclogite". In the first two sentences of the legend of Extended Data Fig. 7, "silica-undersaturated" has been replaced with "silica-saturated" twice. Reference 41, which is cited in the Methods section 'Bulk rock sample preparation', should read: "41. He, Z. et al. A flux-free fusion technique for rapid determination of major and trace elements in silicate rocks by LA-ICP-MS. *Geostand. Geoanal. Res.* **40**, 5–21 (2016)." instead of "41. He, Y. et al. High-precision iron isotope analysis of geological reference materials by high-resolution MC-ICP-MS. *Geostand. Geoanal. Res.* **39**, 341–356 (2015)." . These errors have been corrected in the online version of the Letter.

CAREERS

CAREER PATH The transition to a role outside scientific research **go.nature.com/30qlvh5**

ONLINE Career resources from our scientific community **nature.com/careers**

DATA Advice for managing large data sets **go.nature.com/2lliudb**

TIM WIMBORNE/REUTERS

Mount Stromlo Observatory near Canberra was destroyed by a bush fire in 2003. Although it was devastating for the staff, the calamity led to new initiatives.

LAB SAFETY

When disaster strikes

Researchers whose labs have been destroyed advise flexibility and forward planning.

BY JOSIE GLAUSIUSZ

Glaciologist Martin Sharp at the University of Alberta in Edmonton, Canada, vividly remembers a particular Sunday in early 2017. At midday on 2 April, a freezer malfunction at the Canadian Ice Core Archive (CICA) caused the partial melting of one-eighth of the facility's 1,409-metre collection of ice cores, the oldest of which dated back more than 60,000 years. The centre had opened just six months earlier, and its systems had been functioning reliably.

Not only did the cooling system inside one of the facility's two freezers fail, the warning system that was designed to send an alert in the event of a malfunction failed as well, explains Sharp, who is CICA's principal investigator (PI). The university became aware of the problem only when the freezer temperature hit 35 °C, triggering a fire alarm.

It was a holiday weekend, and all of the scientists were off campus. "When I got there, maybe an hour and a half after it was initially discovered, it was like a sauna," Sharp says, "with steam visible in the air and water on the floor."

Sharp and his colleagues immediately shifted the partially melted cores — some 13% of the collection — into a functioning freezer. In the moment, there was little else to be done, although the team worried about accidentally analysing refrozen meltwater instead of unaffected, pristine ice.

Sharp is among a select club of scientists that few would envy. These researchers' laboratories, or priceless artefacts or samples, have ▶

been damaged or destroyed in explosions, floods, hurricanes, fires or other disasters. Many dealt with the trauma by diving straight into the work of reconstruction. Some report that they emerged from the experience with greater resilience, after introducing better safety standards and installing effective alarm systems, or because they had been given a blank slate to take their research in a different direction. Others derive strength from continuing their research abroad, reasoning that this is the best way to thrive and contribute to their field.

Researchers who have faced a lab disaster say that it is crucial to create safety protocols and to review — and practise — emergency plans regularly with all lab members. Scientists at institutions where hurricanes and other violent weather events occur frequently need to make longer-term plans for coping with climate change, deciding whether to rebuild or relocate.

Connections with other labs — in case it is necessary to move — are crucial. PhD students and postdoctoral researchers can ensure that their labs have suitable safety standards and adhere to them, keeping logs of near misses. They can also protect their work by digitizing it and uploading it to a digital repository, so that a record remains.

Sharp advises PIs to envisage the worst-case scenario and to prepare a protocol, ensuring that staff members know which items are the most crucial for future research and where they are kept, so that these can be rescued first.

Then, he says, the entire lab should run exercises to determine the plan's efficacy. "Don't trust the technology, even if you have a lot of confidence in it," he says, because technology failures are at the root of many lab disasters. And the losses from such disasters can be exacerbated by failures in lab-response procedures.

After the partial meltdown, Sharp's team set about re-evaluating the lab's emergency plan and reinforcing CICA's warning system; twice-daily updates of freezer temperatures are now sent to the mobile phones of ten staff members. The archive has also introduced a system of colour-coded cores, indicating the order in which they should be rescued in an emergency.

TRAUMATIC AFTERMATH

Accidents and disasters can take a huge psychological and emotional toll. Microbiologist Fathiah Zakham has vivid memories of the air strikes that targeted the Red Sea port city of Hodeidah in Yemen in spring 2015, and she still suffers flashbacks from that terrifying time. "We were hearing the voices of explosions, of air strikes, of attacks," she recalls. On 27 May that year, catastrophe struck: a bomb completely destroyed the Faculty of Medicine and Health Sciences at Hodeidah University, where Zakham's lab was located, killing four security men. "It was a very new building, and it became a mass of rubble," she says.

Hurricane Maria flooded Belinda Pastrana's lab in Puerto Rico, prompting her to move to Massachusetts.

PABLO PANTOJA/GETTY

In the aftermath of the attack, Zakham decided to leave Yemen, and won a Swiss Government Excellence Scholarship to do postdoctoral research at Lausanne University Hospital, where she started working in July 2017. She is now a postdoc at the University of Helsinki, where she researches and develops tools for diagnosing viral haemorrhagic fevers.

"I'm working every day at the lab, and I'm writing, researching different articles, I'm attending conferences and different scientific events," says Zakham, who was awarded the 2017 Al-Kharafi Prize by the World Academy of Sciences, which recognizes exceptional female scientists from countries that are lagging scientifically. "But it's very difficult to recover. I still imagine how I was running with my students, how they were crying."

The impact of loss is fresh for anthropologist Luiz Fernando Dias Duarte, who recalls the rage he felt when fire engulfed the National Museum of Brazil in September 2018, destroying its 20-million-strong collection of fossils, books, ceramics, Egyptian sarcophagi and priceless South American archaeological artefacts (see p. 312). "My immediate reaction was of intense anger," says Duarte, the museum's deputy director. The museum, housed in what was once the palace of Brazil's imperial family in Rio de Janeiro, had no sprinkler system and firefighters were ill-equipped: the closest hydrants were broken, forcing them to use water from a nearby lake.

"The risk of fire was very great, and we were completely aware of this," Duarte says, explaining that the museum had planned to move some of its valuable collections to new buildings, with financing approved just three months before the fire. "It was a very hard loss," he adds, explaining that he also lost his own archives, correspondence, books, journals and other publications totalling about 7,000 titles.

But although the blaze took a huge emotional and psychological toll, Duarte rallied quickly, throwing himself into reconstruction, fundraising and seeking replacement collections and donations of books, as well as negotiating new channels for scientific partnership. The museum now has pledges of financing from Germany, the United Nations Educational, Scientific and Cultural Organization and Brazil's federal government, as well as promises of replacement books and specimens from institutions in Europe, China and the United States.

"Go on fighting," Duarte advises scientists who are struggling with such crises. "When I'm engaged in organization, I feel safer."

He and Sharp urge other scientists to be frank about the significance of what is lost in such a debacle. Sharp spent a week responding to what he calls a frenzy of phone calls and e-mails from journalists after the University of Alberta held a press conference on the morning after the meltdown.

That openness, says Sharp, helped a great deal, because media reports reached a larger audience, including those with relevant expertise, who offered to help however they could.

Neither Sharp nor Duarte says that they found media interest in the debacles to be intrusive. "Dealing with media calls did take up a lot of time, but it was time well invested because it created opportunities to talk about ice-core science, climate change and atmospheric-pollution issues that would not have come up otherwise," says Sharp. "It resulted in a lot of positive advice and input from people who had had similar experiences. That has really helped us."

Duarte says that he perceived no negative consequences from his media interviews. It is imperative, he says, that the public continues to be aware of the challenges in the National Museum's reconstruction.

Openness, agrees chemist Dominick Casadonte, is key. Casadonte was chair of the chemistry and biochemistry department at Texas Tech University in Lubbock in January 2010, when then-PhD-student Preston Brown lost three fingers, punctured his eye and suffered burns in a university chemistry lab.

Brown, who was stirring a volatile compound, had removed his protective goggles before the mixture exploded. Casadonte, the third person at the scene, was horrified when he saw the extent of the student's injuries. In an interview two months later with the US Chemical Safety Board, he recalls saying that he never wanted a recurrence in any lab with which he is associated.

Casadonte became determined to improve the safety culture at Texas Tech, and to be open about it. "A lot of schools that have accidents just basically try to cover it up, or circle the wagons, to try to not let things out," he says. "We had a very courageous vice-president for research, Taylor Eighmy, now president of the University of Texas at San Antonio, who said 'No, we're going to be transparent, we're going to be public, and the process of transforming our safety culture can be an example for others around the country.' So we all collectively decided to do that."

New safety rules at the university require all incoming graduate students, staff members and teaching assistants in chemistry to take a course in chemical safety. In addition, they mandate barcoding of all chemicals with safety classifications, and strict penalties, including closure, for labs that ignore safety protocols. Texas Tech also maintains a database of near misses and lessons learnt to help avoid accidents recurring.

BOUNCING BACK

Some scientists say that their institutions emerged stronger after a disaster, with improved facilities. Matthew Colless, director of the Research School of Astronomy and Astrophysics at the Australian National University in Canberra, was a senior fellow at Canberra's Mount Stromlo Observatory when it was destroyed by a bush fire in January 2003. The ferocious fire swept across Canberra, killing 4 people, incinerating more than 400 houses and destroying the observatory's 6 telescopes and all of its workshops.

When Colless surveyed the scene of destruction two days later, he saw only smoking ruins. Fortunately, the modern telescopes — which Colless was using to map dark and ordinary matter in a very large galactic survey — were at Siding Spring Observatory, 600 kilometres away.

Colless left in 2004 to work at the Australian Astronomical Observatory near Sydney; when he returned to Mount Stromlo as director in 2013, he says, the observatory was in a much stronger position than it would have been if the fire had not happened. Observatory directors were able to reconstruct and rebuild with financing from the Australian federal government — after a long battle with insurers.

Now, because the observatory no longer has to maintain heritage telescopes — the oldest of which dated back to 1911 — it can invest in new initiatives. "Once they're burnt to the ground, they don't require a lot of money to maintain them," Colless says. "You put a plaque up in front of them, let the rain in, and let them become romantic ruins."

"It's very difficult to recover. I still imagine how I was running with my students, how they were crying."

Like Sharp, Colless advises developing a plan for when disaster strikes — including good insurance. But that plan, he adds, does not necessarily mean rebuild. "You have to stop and think, is that really what you want to do? Maybe there are other, smarter things you can do. Every one of these crises is an opportunity."

Long-term planning might be crucial for universities affected by more frequent flooding caused by climate change, says Charles Connerly, director of the School of Urban and Regional Planning at the University of Iowa in Iowa City.

That university, which is built on a flood plain, suffered an estimated US$743 million in damage during a flood in June 2008. The university relocated some departments and rebuilt them on higher ground, with the aid of federal funding.

Connerly argues for a regional solution, despite its cost: to restore the entire watershed to a more natural state that could better absorb flooding, by building and restoring wetlands. "We don't know what climate change is going to bring. The amount of storms we're getting is increasing dramatically," he says. "If it's only going to get worse, then maybe we have to come up with a more appropriate solution, one that better respects the watershed."

BLESSING IN DISGUISE

Kathryn Moore, whose work at New York University (NYU) focuses on the immunology of cardiovascular disease, also recommends recalibrating following a disaster. In October 2012, her basement mouse facility in New York City flooded in the wake of Superstorm Sandy, and the simultaneous power cut caused the loss of hundreds of NYU's unique mouse strains, including ten of Moore's, which had taken years to develop. She also lost many of her tissue sections of atherosclerotic plaques that were stored in −80 °C freezers.

"Losing all that was almost like starting from scratch," Moore says. "It was an amazing experience of being given a blank slate: to take the most exciting projects and think about what we really wanted to do. It changed the direction of my science." She and her team moved away from studying atherosclerosis in live mice to short-term *in vitro* studies on non-coding RNA.

"I think that rather than spending time feeling sorry for ourselves, we focused on how we're going to rebuild. That was really important for morale," Moore says. She recalls that individual responses to such a disaster can vary widely. Although some respond with action, others in her team of 15 scientists were traumatized and had a difficult time moving forward. "I learnt not to judge people who got stuck in place," she says, "and to be patient." Moore was able to place two of her postdocs in colleagues' labs elsewhere.

Belinda Pastrana also dealt with relocation after Hurricane Maria struck Puerto Rico in September 2017. She moved her company, Protein Dynamic Solutions, from an incubator site associated with the University of Puerto Rico to Boston, Massachusetts. Pastrana, the business's founder and chief executive, and her team had developed a laser infrared microscope to evaluate protein structure and stability.

Maria destroyed the Caribbean island's power grid and blew away some of Pastrana's solar panels. She lost 20 years of work, and her company's waterlogged lab, invaded by mould, was rendered unsafe for work.

Had she stayed at the university where she had taught for 20 years and helped to create a PhD programme, she would have had to abandon her business and its important technology. "I felt a moral responsibility to make this company a success," she says. By continuing her research elsewhere, she can provide funding to the university through royalties from her licensed patents.

Pastrana recommends that researchers continue to build and foster their scientist networks long before any crisis, as well as afterwards. She found alternative labs for her graduate students to complete their PhDs. Think outside the box, she adds, and "find alternate routes to pursue your science and your passion".

That, too, was Zakham's strategy. "Moving from the country [Yemen] was not an option — it was an obligation," she says. In leaving, she was able to reignite her career, and support female scientists back home. "Life should continue," she says. "We shouldn't give up." ■

Josie Glausiusz *is a freelance writer in Israel.*

SHIPMASTER'S SCALP

How to buy time.

BY JEREMY SZAL

They come to see me on day 289 of my captivity.

My throbbing head feels stuffed with metal wool, but I smear a smile on my face as Shipmaster Hargreave and Detention Officer Bossa unlock my cell. I'm strapped into this cradle by my arms, legs, wrists, waist, chest, thighs and neck, a thick prisoner's harness wired tight around my torso and magnetized to the rear of the cradle. And that's *before* they sealed me inside my prisoner's exosuit.

All in all, I take it as a compliment.

Bossa checks the status on a holograph. My smile widens as she glances at Hargreave. "No progress."

Hargreave raps her knuckles against the Wiring embedded in my temple. "Oh, he'll break, sooner or later. Or, you could just tell us. We'll dig it out eventually, Kharrus."

"Well," I rasp through the steel mesh muzzle affixed to my face, "you've done a great job so far."

Bossa's fist smashes into the side of my head. I roll with the blow, spit bloody saliva. I've had worse in spacedock bars.

"I've got to say, being a smuggler while Wired is a pretty stupid career move."

For once, Hargreave has a point. When implants, or Wires, came around, everyone thought they'd be exclusively purchased by the wealthy. The opposite was true. Wires became so cheaply manufactured that anyone hoping to maintain pace with modern life — storing, backing up and accessing memories and intel across the datasphere instantaneously — had to get one. On some planets, it's compulsory. Rich folk can afford remote storage accessed by internal wetware, the kind that isn't hackable and doesn't have their memories and places they've been and what they did registered in a semi-public database. In effect, you could afford *not* to be Wired.

It's partially why I turned to smuggling. People across the Systems will pay solid u-credits for materials they don't want to register with customs. Narcotics, booze, databanks, artefacts, military-grade weapons, relics. Anything. Me and my crew had a solid career going for about nine standard years. I made one slip-up, but one is all it takes. I was captured by Systems Security and dumped here in a deep-systems Detention Centre for interrogation.

Hargreave leans against the mirror-smooth wall. "This could all be over if you gave up your crew."

I snort. "You don't do this often, do you?"

"Hey, we've got time. You don't."

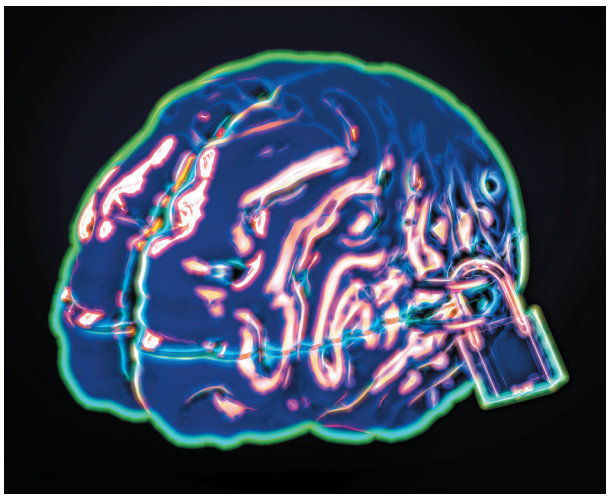


ILLUSTRATION BY JACEY

For nearly 300 days I've sat here, resisting the Scalper software they've fed into my Wire as it sniffs through my mental server cabinets. Trawling for jobs I've pulled, clients I work for, items of interest I've smuggled. And most importantly: where my crew is going. My head pulses with a dull ache as I combat the Scalper. Filling my head with distractions and false memories and random statistics as if they were true, confusing it. Killing research patterns, mentally burying data. It's a literal battle of wills. An AI can't tell which memories are legitimate and which are fabricated, not unless they want to turn my brain into a stack of smoking neurons. But it's adapting, recognizing patterns, getting to know how I think. Perhaps I could fight it, if my exosuit weren't limiting my sleep to four hours a night, lowering my food and water intake to minimum. Blasting me with relentless white noise while wrapping me in sweltering, sauna-like heat, or borderline-permafrost colds. Civilized torture, barely within intergalactic legalities. My body is numb. I can scarcely stay awake, let alone perform psychological warfare.

But every day I resist, I give my crew one extra day's head start. That's how we dealt with botched broker deals. Not focusing on surviving next week, next month. Just tomorrow. We could effectively postpone problems forever, as long as we lasted until the next day. Now, I'm doing that for my crew, day by agonizing day. It's what any good Shipmaster would do, and I pride myself on being the best. We swore absolute loyalty to each other when we became smugglers. A man's only as good as his word. I won't let it be said Alistair Kharrus lasted to anything less than the absolute breaking point.

"I hope he doesn't talk." I glance over my shoulder where Bossa's fiddling with my exosuit's restraints, tightening them with bone-crushing force. "More fun that way."

Hargreave leans in close enough to kiss, flicks a tattooed finger on my Wire. "You're a businessman, Kharrus. So let's do business. You talk now, I'll release you from the exosuit, unplug the Scalper. Give you a first-class stateroom. Hell, finger your crew in the line-up and I'll get you a Cobalt-class mental substrate. Not that cheap backalley trash. What do you think?"

I let her know by headbutting as hard as I can. Bone crunches and blood sprays. She staggers back, holding her broken nose. Despite my weariness, I wear a face-splitting grin behind my muzzle. "Is that really the best you can do?"

Bossa's about to shatter my jaw into sugar-glass when Hargreave stops her. "No. Crank the Scalper up to the next level. Actually, make it three levels. Get it to dig deep, permanent damage be damned. See how rebellious he is after a few more months. Kharrus, when you're drooling and being fed through a tube, you'll wish you took the offer."

Bossa overrides the system to adjust the settings below legal minimums. Machinery whirls and cranks. I grin at them as they leave me to prepare for the mental battle to come. My headache turns cold and throbbing as the Scalper burrows savagely through the geometries of my brain. My restrained fists shake as I encase the names of my crew in mental concrete. Preparing to resist for one more day. And the one after that. And the one after that.

That's all I need. Just one more day. ■

Jeremy Szal *writes about galactic nightmares, wide-screen futures and characters fighting for hope in dark worlds. He is author of the dark space-opera novel* Stormblood, *forthcoming from Gollancz in February 2020. Find him at jeremyszal.com or @JeremySzal.*